

Who will fashion the distant future?

Two large companies, IBM in the United States and ICI in Britain, suddenly look vulnerable. Is the explanation their sheer size or, more probably, their ingrained habits?

INTERNATIONAL Business Machines (otherwise IBM) is one of the largest companies in the world. Its annual turnover is greater than the gross domestic product of many substantial nation states, Belgium for example. Apple Computer, by contrast, is one of the youngest and, according to legend, began life in a garage just over a decade ago. Why should these two very different companies be planning a joint venture that, they hope, will fashion for us computers of a radically different kind a few years from now (page 98)? And will they succeed? What, for that matter, explains the present fuss in Britain, where Imperial Chemical Industries (ICI), one of the largest indigenous companies and one of the most technically advanced, is behaving like a rabbit transfixed in the face of the threat implied by the purchase of nine per cent of its shares by Hanson Industries, a largely financial organization that has made its successful way in the world by buying other companies and selling their bits and pieces separately?

There is a sense in which both IBM and ICI are victims of their own success. Until a decade ago, when its first personal computers appeared in the shops, IBM was the world's predominant supplier of mainframe computers. It had won a superb reputation for keeping this complicated machinery at work, but, perhaps for that reason, it was slow to recognize how far the personal computer would change the world. Then, having become the chief provider of personal computers worldwide, it was slow to appreciate that its customers would come to demand ever greater amounts of immediate and long-term memory, intellectually more transparent computer chips and machines on people's desks that would match the power of its own chief products — mainframes hallowed over years. So, last breathtaking week, IBM has jumped into three entirely novel joint ventures — an arrangement that Wang Laboratories should sell chiefly IBM computers, joint European production of memory chips with Siemens and the Apple venture (Department of Justice allowing).

ICI's problem is similar in origin, but the solution is not immediately apparent. Although British-based, ICI is international in scope. Also like IBM, the thoroughness of ICI's technical backing for its products has won it an enviable reputation. But, like all conglomerates, ICI's financial performance is necessarily an average of the performance of its divisions, some of which are pretty unglamorous in the middle of a recession. For several

years, ICI has been trying to back out of well-established ventures that add relatively little value to the cost of the raw materials they employ, but without enough success. The high-value end of the business, on the other hand, is in pharmaceuticals, which is a notoriously high-risk business. A decision to put all its eggs in that basket would not have won the company many new friends.

So does it follow that all big companies are as vulnerable as IBM (in a quickly changing market) and ICI (hurt by the recession)? Not necessarily. If the threat to ICI is that Hanson might buy the whole company (at about £12,000 million at this week's stock market price) and then sell the pieces at a profit, the company itself could in principle do the same, providing a similar profit for its shareholders. The casualties of that process would be the corporate culture (which is unimportant) and the common services that the various divisions enjoy, of which research and development is the most important. It would be a great misfortune for ICI's separate enterprises, and for Britain as a whole, if its substantial research and development effort were sacrificed on the altar of immediate profit as a successful Hanson bid would almost certainly entail. The other side of that coin is that big companies such as this, which have the opportunity, also have a responsibility to see that their research does even better than that of smaller companies. Despite its long record in the field, that is where ICI should have done better in the past fifteen years.

IBM's difficulty is different in kind. The sheer physical inertia that has prevented it from responding as quickly as it might have done to novel challenges is only part of the explanation. Over the past fifteen years or so, it has been compelled to make choices between different paths of development for fear of creating circumstances in which different parts of the same company were competing with each other. To have been selling cheap high-capacity machines when Sun Microsystems began (in the early 1980s) would have seemed a corporate nonsense, as it would have been to have embraced both the UNIX operating system and IBM's own version of the DOS based systems. But would that necessarily have been the case? Would motor-car manufacturers equipped to make and sell invalid carriages be thought foolish if they also sold sports cars? Backing all possible horses would have been a better bet for such a giant than the small-company trait of making risky choices. □



Human genome databases at the crossroads

- Learning to cope with avalanche of data
- Will biologists accept rapid release?

Washington

As the Human Genome Project gets fully up to speed, genome researchers face a problem: information will soon be accumulating so fast that the prompt sharing of gene-mapping data threatens to become a bottleneck for the planned 15-year genome initiative. Data-handling experts and software engineers who design new computer databases say that action is needed now to handle the avalanche of data.

These concerns will be raised at a meeting later this month of the directors of the National Institutes of Health (NIH) genome centres. But the solution to the problem may be hampered by a clash of visions between the biologists working on the genome project and the computer scientists designing ways to disseminate rapidly the gene-mapping data. The computer scientists' proposals are forcing the biologists to determine their position on a fundamental, but still unresolved, issue: how quickly are they willing to let their data be made publicly available?

The main repository for human gene mapping data is now the Genome Database, run from Johns Hopkins University in Baltimore, Maryland. It contains a 'consensus map', in which all the information has been carefully screened, to iron out any inconsistencies in data submitted by different researchers.

But as the volume of raw data coming into Hopkins increases, the delay between submission and the appearance of screened data in the database will increase. "People are going to have to give up on consensus data if they are to have rapid access to gene-mapping information," says Nat Goodman, an information scientist at the Whitehead Institute of the Massachusetts Institute of Technology.

Within two years, Goodman predicts, the consensus process will be overwhelmed. This makes it essential, he says, that a new system of intermediate databases be developed to act as clearing houses for unscreened data.

Many information scientists agree. Peter Pearson, who runs the Genome Database at Hopkins, thinks that these intermediate databases should be the responsibility of the individual genome centres set up by NIH and the Department of Energy. Each centre is the hub for work on one or several individual chromosomes, and is supposed to act as a resource for investigators working on these chromosomes at other laboratories.

Although few of the main centres have

made significant progress on the informatics front, database experts are hard at work at these centres, and if all goes as planned, should have intermediate databases ready within a couple of years.

That will still leave the problem of communication between the different databases, which may have different formats. Researchers expect that it will soon be possible to design a 'mediator program' allowing researchers to communicate with each of the major databases via the NSFNET computer network.

In the meantime, Goodman and Thomas Marr from the Cold Spring Harbor Laboratory would like to set up a new central database to handle mapping data before it is screened for inclusion in the Genome Database. This could be set up in a year or so, for as little as \$100,000, Goodman says.

Goodman and Marr's vision, however, may get a cold shoulder from some researchers. Elbert Branscomb, of the Lawrence Livermore National Laboratory, recalls the stunned reactions of biologists at a seminar last year, when he explained the potential for rapid dissemination of data from Livermore's chromosome-19 database. The attitude of some researchers was, he says: "If you do that, we won't collaborate with you any more."

Besides an understandable reluctance to share data with directly competing research groups without at least some delay, many biologists are worried about the potential embarrassment caused by the general distribution of gene-mapping data that have not been fully checked for accuracy.

Leroy Hood, of the California Institute of Technology, believes that controls on data access should be sensitive to the confidence limits placed by researchers on the accuracy of their own data. When data are ascribed a confidence level of only 80 per cent, for example, Hood believes that they should be distributed only to biologists for whom the information is directly relevant in their own research.

Working out an acceptable data access policy for the next generation of genome databases will, however, require a greater degree of communication between the 'two cultures' of scientists working on the genome project than has so far been the case. At most genome project meetings, Hood says, "informaticists speak to informaticists and biologists speak to biologists."

Peter Aldhous

NEWS IN BRIEF

US okays Antarctic pact

Washington

A potential rift among the Antarctic Treaty nations has been averted, now that the US Administration has agreed to sign a protocol to protect the continent's environment. Last month, the US delegation to a meeting in Madrid irritated other delegates by refusing to agree to a version of the protocol that had been revised specifically to meet US concerns. The controversy was over a provision prohibiting mining or oil exploration in the continent for at least 50 years.

The version now accepted by the United States allows the Antarctic nations to amend the mining ban after 50 years. If three-quarters of the current 26 parties to the treaty support a relaxation of the ban, but a minority of nations refuse to ratify the change, any country can withdraw from the protocol after another two years.

P.A.

Hubble fix on course

Washington

National Aeronautics and Space Administration (NASA) officials begin negotiations this week with contractors hoping to build the COSTAR optical correction device for the Hubble Space Telescope. COSTAR will place a series of lenses in front of three of the instruments at the base of the telescope, in the hope of correcting the spherical aberration of Hubble's primary mirror.

To accommodate the \$20-million COSTAR, one instrument — the High Speed Photometer — must be removed from the base of the telescope. NASA hopes to have completed most of the interesting scientific research with this instrument by the time the fix is attempted, during a 1993 space shuttle flight.

P.A.

Stanford unrepentant

Washington

In a move that seems likely to prompt further outrage in Congress, Stanford University has told federal officials that the university should be reimbursed for indirect research costs to 1991-92 at the rate of 76 per cent of the value of its research grants. This is 2 per cent more than the rate that sparked a congressional investigation earlier this year and precipitated the current indirect costs imbroglio.

"We know the political winds are shifting [and] . . . expect that this rate will be revised," says Byers, but he maintains that the starting point for negotiations must be Stanford's estimate of its true research costs. In calculating the overhead rate, Stanford has not taken into account the 26 per cent limit on administrative costs proposed recently by the White House Office of Management and Budget. Stanford was given a temporary cut from 74 to 55 per cent earlier this year.

P.A.

An 'alternative' in Germany

Freiburg, Germany

A new \$69 million cancer research centre now being built in Freiburg could help bring clinical cancer research in Germany out of its current doldrums. The Tumour Biology Centre, which is expected to open its doors in early 1993, is the first large-scale attempt in Germany to wed clinical cancer medicine with basic science — an approach that has gained wide favour in the United States and Britain and that has been urged by German research organizations for years, but to limited effect.

Despite its promise, however, there are serious doubts about whether the centre can succeed in its current form. It owes its existence to a promise by its director to study 'alternative' cancer treatments, and German researchers not affiliated with the centre question how easily it can attract serious scientists if this promise is kept.

The founding director, Gerhard Nagel, a respected oncologist who gave up a chair at the University of Göttingen to start the centre, raised funds from a variety of sources including three pharmaceutical companies and a bank.

But the contribution that has caused the most comment among other researchers is the DM25 million from the *Land* of Baden-Württemberg, which has also agreed to provide DM2 million a year for operating costs. Nagel specifically promised politicians from Baden-Württemberg that the centre would look into unconventional treatments — a promise he intends to keep. "We will take alternative medicine seriously and put it to the test," he says.

But it will be a tricky task for Nagel to keep his promise without alienating mainstream researchers. "If [Nagel] starts to test herbs and juices on people [to the exclusion of real science], he will lose his reputation overnight," says Sabine von Kleist, director of the institute for immune biology at the University of Freiburg. Von Kleist's view has been echoed by cancer researchers across Germany.

Nagel has assured colleagues that he will use only scientific methods in studying both conventional and unorthodox treatments. "We will not touch treatments that cannot be tested scientifically," he asserts.

More than just the future of the centre depends on how well Nagel is able to balance the conflicting demands placed on him. If it succeeds, the new centre — which will focus on tumours of endocrine origin and will include both a 200-bed clinic and a research branch concentrating on clinical pharmacology — offers a way to cast off the institutional shackles that have prevented Germany from doing more cutting-edge research in clinical medicine.

University politics has traditionally been the biggest hurdle to achieving better clinical research in German hospitals. Because of the way clinical research is organized at universities, physicians are usually overwhelmed with responsibilities for teaching and patient care which leave them little time to do research.

Furthermore, clinic directors are entitled to share in the clinic's profits from private patients, so they have little incentive to reinvest those profits in research.

The Nagel centre manages to overcome both these disadvantages. Staff members at the Freiburg centre will not have teaching responsibilities (although they may teach if they choose), and the profits from the clinic will all be put back into the research effort, Nagel says.

Nagel's centre is also innovative in its financing scheme, which brings in outside investors. He has offered his three pharmaceutical backers — Asta Pharma of Frankfurt, Schering of Berlin and Ciba-Geigy of Basle — first right of refusal on any new compounds to emerge from the centre. Although this funding mechanism has existed for years in the United States,

it is unheard of in Germany, where state-owned university hospitals are not allowed to make such arrangements.

A further source of financing may turn out to be the producers of the so-called 'alternative' cancer treatments, which range from extracts of mistletoe and extracts from the thymus glands of animals to 'fresh-cell therapies' which involve the injection of sheep cells into the body. The field of 'alternative medicine' in general, which includes homoeopathy and herbal medicine, commands a huge market in Germany, where almost every pharmacy also dispenses homoeopathic remedies. If Nagel manages to stay on the good side of the producers of alternative treatments, they may provide a rich untapped market for the financing of further clinical trials.

In the end, the future of the centre will depend on the quality of the scientific staff it can attract, and that in turn will depend very much on Nagel, a respected but rather mercurial character of whom other researchers tend to be wary, if not downright suspicious. Privately, they say they wonder if picking a path for the new clinic that avoids the many pitfalls may be too much for even him.

Steven Dickman

Barking up the right tree

THIS photo does not depict the work of giant squirrels, but rather the damage caused by a band of Oregon poachers harvesting bark from Pacific yew trees *Taxus brevifolia*. The bark contains the drug taxol,



which is showing great promise in clinical trials as an anti-cancer agent, creating a burgeoning demand for the bark.

The US Forest Service, which oversees the national forests in the northwestern United States where many of the yew trees are found, is offering a \$10,000 reward for information that helps to catch the poachers. The service signed an agreement last month with the drug company Bristol-Myers Squibb to

supply yew bark for taxol extraction, and spokesman Jim Sanders says all harvesting must be coordinated through the new agreement. Bark poaching has so far been a small problem — with the number of trees attacked only in the hundreds — but the Forest Service wants to ensure that it does not get out of hand.

Over the coming year, the National Cancer Institute hopes to treat 12,000 cancer patients with taxol, either as part of the continuing clinical trials or for 'compassionate release' of the drug to severely ill patients, says Jason Fisherman from the institute's Investigational Drug Branch. This will require up to 40,000 trees, from the approximately 23 million in the states of Washington and Oregon.

Eventually, taxol may be produced without the need to sacrifice yew trees. Saul Schepartz, from the cancer institute's Developmental Therapeutics Programme, expects taxol extraction from yew needles to become significant within three years. And if experimental cell-culture techniques can be developed on a commercial scale, the demand for the bark may be over in five years.

Taxol fights tumours by inhibiting the formation of normal mitotic spindles, thus reducing cell proliferation. P.A.

Japanese set to X-ray Sun

Tokyo

NEXT month, Japan will launch yet another X-ray observation satellite, this one designed to observe solar flare activity of the Sun. This series of satellites has allowed Japanese astrophysicists to capture a leading position in the field of X-ray astronomy.

The Solar-A satellite, due to be launched on 26 August, will be the world's only scientific satellite dedicated to observation of the Sun during the present period of maximum solar flare activity. It will carry both hard and soft X-ray telescopes, as well as a Bragg crystal high-resolution X-ray spectrometer and a wide-

band spectrometer for detecting soft X-rays, hard X-rays and gamma rays.

The soft X-ray telescope, developed in collaboration with US scientists, will be able to detect and take automatic snapshots of the thermal activity of solar flares at intervals of a few seconds.

Meanwhile, the hard X-ray telescope will pick up the activity of high-energy electrons in flares with a time resolution of a second or better. Solar-A follows the institute's Hinotori satellite launched in 1981, which observed solar flare activity during the last solar maximum and helped to establish Japan's foothold in this field.

In the intervening period of quiet solar activity, the Institute of Space and Astronautical Science has developed its expertise in X-ray satellite observations by launching the X-ray astronomy satellites Tenma (in 1983) and Ginga (in 1987). In 1993, the institute will launch a follow-up satellite to Ginga called ASTRO-D.

This series of satellites has nurtured the growth of Japanese X-ray astronomy considerably. Indeed, more than 50 per cent of 160 authors contributing papers on X-ray astronomy to the *Astrophysical Journal* in 1990 were either Japanese (25 per cent) or Western scientists who used Japan's Ginga satellite.

A key reason for Japan's success has been the ability of the space institute to launch low-cost satellites (a few tens of millions of dollars) every few years, and thereby maintain almost continuous observations and hold together and build teams of researchers. In contrast, satellites in Europe and the United States tend to cost hundreds of millions of dollars; the result is long gaps between satellite launches because of lack of funds, making it impossible to collect data continuously and very hard to keep research teams together.

The Japanese space institute can produce low-cost satellites because the institute's scientists and engineers themselves design and build the satellites, launch them, collect the data and publish the results. There is only minimal use of outside contractors.

This approach, however, does have its disadvantages. The US National Aeronautics and Space Administration (NASA) and the US Jet Propulsion Laboratory recently suggested that the Japanese space institute should use NASA ground stations in several places around the world so that Solar-A can maintain continuous observations. But the institute's Yoshiaki Ogawara points out that as the institute's scientists have to do everything themselves, it would be "almost beyond their capability" to man the operations desk 24 hours a day.

David Swinbanks

Erato announces new research targets

Tokyo

JAPAN'S Research and Development Corporation has opted to support a mixed bag of projects this year under its unique Exploratory Research for Advanced Technology (ERATO) programme. ERATO, which has a history of heralding new lines of research that are being targeted by Japan.

The latest ERATO awards will support research on man-made π electron materials, artificial molecular catalysts, bio-fouling and cell switching.

The ERATO programme, which won high praise in a study a few years ago by the US National Science Foundation (see *Nature* 337, 196; 1989), targets research with no immediate applications but that seems likely to produce applications in a few years. Each ERATO project provides a team of 15 to 20 young researchers (all under 35 except for the project leader) from industry and academic institutions with funds of \$2-\$3-million a year for five years — an enormous amount of money by Japanese government standards.

The fields chosen often give an indication of where Japanese research and development is heading. For example, ERATO has supported several projects in nanotechnology in the past few years, and industry and the Ministry of International Trade and Industry are now planning to launch a major national project in this area (see *Nature* 351, 90; 1991).

Four projects will begin this fiscal year. A team led by Susumu Yoshimura, director of Matsushita Research Institute in Tokyo, will try to find man-made materials which — like naturally occurring substances such as graphite and molecules involved in photosynthesis — carry pi-orbital electrons that can wander over large distances on the atomic or molecular scale and thus give unusual optical and/or electronic properties to materials.

Ryoji Noyori of Nagoya University's department of chemistry will lead a team that hopes to create new organometallic catalysts for such purposes as catalysing the formation of molecules or polymers that are specifically either left- or right-handed. Such materials, called optical isomers, often exhibit unusual optical, electrical or medical properties when produced with a pure handedness.

Other ERATO awards go to Hiroto Okayama of the Research Institute for Microbial Diseases in Osaka University, who will search for human genes which control the initial step of cell division and thus play a key role in cell growth and cancer development, and to a team led by Nebuhiro Fusekani at Tokyo University which will work on biofouling.

David Swinbanks

WEAPONS TRADE

Japan embarrassed by weapons sale to Iran

Tokyo

IN an untimely embarrassment for the Japanese government, police last week raided a subsidiary of a major Japanese electronics company that is suspected of exporting high-technology military related equipment to Iran during the Iran-Iraq war.

Disclosure of the exports by Japan Aviation Electronics, a subsidiary of NEC, comes at a time when Prime Minister Toshiki Kaifu was planning to portray Japan at this month's summit of leading industrialized nations in London as a model country that does not engage in military-related trade.

The executive managing director of Japan Aviation Electronics admitted on 5 July that the company had repaired crucial components of the guidance system of US-made Sidewinder missiles for Iran during the Iran-Iraq war. Police also suspect that the company manufactured gyroscopes and other components of the inertial navigation systems for missiles and aircraft and exported them to Iran at a time when the United States and other Western countries had placed an arms embargo on the Middle East country.

The exports are a clear violation of Japan's strict ban on military exports. And government officials say that, if the charges are substantiated, the company could be prohibited from engaging in the export business for three years under regulations introduced following the Toshiba scandal of a few years ago. In 1987, a Toshiba subsidiary was found to have exported sophisticated milling machines that allowed the Soviet Union to make silent propellers for submarines.

Foreign Ministry officials argue that the government's harsh response to the affair indicates that Japan is serious about arms trade control.

D.S.

Weather satellite woes

Washington

WITH two of its three major satellite programmes in trouble, the US National Oceanic and Atmospheric Administration (NOAA) is facing an unpleasant predicament in the near future.

If the satellites do not get back on track, NOAA could find it difficult to know how to provide weather and other remote sensing data over the next few years. Without those data, US weather forecasters could find themselves short of the necessary information to make their predictions, and environmental scientists who depend on Landsat satellites to study vegetation could find their research stalled.

Among NOAA's satellite programmes, the next generation of US geostationary weather satellites is badly behind schedule, way over budget and not even working properly. Meanwhile, the planned seventh satellite in the Landsat series is threatened by a lack of funds. Only NOAA's polar orbiting weather satellite programme is not a matter for concern.

The most immediate concern centres on the GOES-NEXT series of five geostationary weather satellites. Unless the National Aeronautics and Space Administration (NASA), which hired contractors to build the satellites, can convince NOAA officials by the end of this month that last-minute design faults will not further delay the programme, NOAA may be forced to find alternative ways to gather data for use by the US National Weather Service.

The first GOES-NEXT satellite was originally due for launch in 1989, but the most optimistic launch date now is late 1992. The delays have caused the programme's budget to spiral from \$578 million to more than \$1,100 million — assuming that there are no further difficulties. That seems unlikely, however, now that NASA has informed NOAA officials that space-simulation tests in a vacuum and at low temperature have revealed faults in the detectors and circuitry of the satellites' imaging system. "There is a credibility problem" with NASA's optimistic launch schedule, says Thomas Pyke, NOAA assistant administrator in charge of satellites and information services.

If GOES-NEXT is delayed too much, it could leave the United States without adequate satellite coverage for weather forecasting. Since the imager on the GOES-6 failed in 1989, there has been only one geostationary satellite, GOES-7, hovering over the United States. Pyke says that GOES-7 should continue giving reasonable data until 1995, and that the European Meteosat-3 will be moved westward in August to give better coverage of weather systems brewing in the Atlantic.

But any further slippage in the GOES-NEXT schedule may force NOAA to look elsewhere for its next geostationary weather satellite, he says. Purchasing a satellite from European manufacturers, or simply ordering a replica of the existing GOES-7 are among the options being considered.

Pyke admits that the present problems stem in part from the decision, taken by jointly by NOAA and NASA in the early 1980s, to build the GOES-NEXT satellites to an untried specification. (The satellites are three times larger than their predecessors, equipped with imagers covering a broader range of wavelengths.) Dan Tarpley, from NOAA's satellite research laboratory, observes that the delays and instrument problems would have been less serious for an experimental NASA satellite, but NOAA's operational weather satellites "have to work, and have to go up on time".

Compared to the GOES-NEXT fiasco, the long-established Landsat terrestrial imaging programme is superficially healthy. Landsats 4 and 5 are both still functioning, and the sixth satellite in the series is scheduled for launch in July 1992. But the programme's future management and continued funding is in doubt.

Shortly after the Landsat programme was transferred from NASA to NOAA in 1983, then-President Ronald Reagan decreed that Landsat should be a commercial operation, eventually financing itself through the sale of its images, which retail at several thousand dollars a frame. The anticipated market, however, has failed to materialize, and despite assurances from the US Administration that Landsat 7 will be funded, so far nothing has appeared in a presidential budget request. In an attempt to prevent a gap in Landsat imaging after the sixth satellite fails around 1997, Representative George Brown (Democrat, California), chairman of the House Science, Space and Technology Committee, launched a bill in May to authorize funds from the 1992 budget to begin work on Landsat 7. When the 1992 budget appropriations bill left the House, \$30 million was earmarked for Landsat 7, and Brown's staff hope that more may be added in Senate in the coming weeks.

But if Landsat is to have a sound financial future, it may be necessary to revamp thoroughly the programme's management. Many observers see NOAA as a strange home for Landsat, because the terrestrial images produced by the satellites have little connection with NOAA's responsibility for the oceans and atmosphere. Both Brown's committee and the National Space Council are looking at the possibility of setting up a joint programme office run and financed by all

the agencies that are interested in Landsat data.

Securing the long-term future for the Landsat programme would be easier if the diverse Landsat data user community could agree on a preferred specification for Landsat 7's instruments. The present Landsat satellites can resolve images down to 30 metres at ground level, but William James, director of the Defense Mapping Agency, told a joint hearing of the House Committees on Science and Intelligence last month that the military could make better use of the data if Landsat images had a resolution of less than 5 metres.

Such fine resolution would actually hamper the work of environmental scientists using Landsat images to study vegetation cover. Barry Rock of the University of New Hampshire said at the same hearing that his studies of acid rain damage to Eastern European forests would be impossible with 5-metre resolution. Patches of damaged forests would then be indistinguishable from shadowing in the canopy, he said. **Peter Aldhouse**

INDIA

Sterilized dogs, bulls will keep their libido

New Delhi

INDIA plans to sterilize stray dogs and 'scrub' bulls with an animal birth control vaccine developed at the National Institute of Immunology in New Delhi.

A single injection makes male animals permanently aspermic, with the sperm count in semen dropping to zero within four weeks, the institute claims. The sterility produced is permanent and the animal does not lose libido. The cost of sterilizing the animal is said to be less than one rupee.

The drug controller of India has given "new drug authorization" for this vaccine after extensive trials for efficacy, safety acceptability in bulls, buffaloes, dogs and rams. It will be marketed under the trade name Talsur and has been licensed to a public sector company in Bangalore.

According to institute director G. P. Talwar, the Talsur vaccine will help the rabies control programme as well as the current campaign to boost milk production in cows through artificial insemination. Stray dogs will receive a single injection to stop their proliferation, ending the current inhumane practice of capturing and killing them to prevent rabies. Sterilizing scrub bulls — wild bulls with usually inferior genetic characteristics — should improve the artificial insemination programme by ending the current problem of scrub bulls impregnating cows before arrival of semen from semen banks. **K. S. Jayaraman**

IBM and Apple set up camp

Washington

In one of the most important developments yet in the consolidation of the US computer industry into a handful of competing camps, the two main US personal computer manufacturers, IBM and Apple, last week signed a letter of intent to cooperate in developing hardware and software for the next generation of desktop computers.

If successful, the alliance promises to provide computer users with faster desktop machines, capable of doing many tasks at a time. Although other consortia are also working to this end, the competitive spur provided by the new link-up may accelerate the appearance on the market of faster machines. It should also become possible to use Apple's Macintosh machines as terminals to work on IBM mainframes, producing easier-to-use computing networks for the many computer users who have grown accustomed to Apple's graphically based systems.

With the costs of research and development in the computer industry forever rising, and the profit margins of the major companies being squeezed by advent of smaller companies selling cheap 'clones' of the leading designs, the formation of strategic alliances has become a general trend.

The most significant feature of the wide-ranging IBM and Apple technology agreement is the announcement that the two companies will work together to develop a new personal computer operating system (the software that controls a computer's basic functions), using object

oriented programming techniques. This operating system will be designed to run on a range of different machines.

The agreement's other main component is the decision to further refine IBM's state-of-the-art reduced instruction set computing (RISC) microprocessor, one of the fastest microchip designs available. Motorola will develop more compact and cheaper versions of IBM's RISC chip, for use in future IBM and Apple computers. Greater computing speed may allow future desktop computers to run programs combining text, graphics and video — so-called 'multimedia' computing.

While most industry analysts agree that successful collaboration between IBM and Apple would redefine the landscape of the US computer industry, there is still some scepticism about the likely success of the arrangement.

Contracts to formalize the collaborations described in the letter of intent have yet to be thrashed out, and several much-heralded computer industry alliances signed in the past have not yielded significant results.

The new IBM and Apple link may pose more problems than most prospective alliances, simply because of two companies' differing business styles. IBM has traditionally been seen as a rather staid organization no bad thing when a company's principal market is the conservative world of corporate America. By contrast, the California-based Apple has presented itself as a pioneering, freespirted company, ever ready to take risks.

Peter Aldhous

A European collaboration for IBM

INTERNATIONAL Business Machines (IBM) and Siemens of Germany will join forces to manufacture next generation computer memory chips, a move aimed at challenging Japanese domination in that market. The news came on 4 July, just one day after IBM announced a major collaboration with Apple Computer to develop personal computers and work stations (see above).

The IBM-Siemens agreement is likely to alter the balance of the chip-making industry worldwide, and particularly in Europe. It ends the chance of a merger or close collaboration between Europe's three major semiconductor manufacturers, Siemens, Philips Electronics of the Netherlands and SGS-Thomson Microelectronics of France and Italy — a strategy urged by some as a way to create a purely European semicon-

ductor concern able to compete on equal terms with US and Japanese companies. But because IBM and Siemens are leaving the door open for other partners, their agreement may allow European chipmakers access to state-of-the-art technology and could eventually increase Europe's competitiveness in the semiconductor field.

IBM and Siemens will expand an existing IBM semiconductor plant in France to produce 16-megabit dynamic random access memory chips, or DRAMs. The two companies will share the \$700 million cost of construction. The chips will be made according to an IBM design and with IBM manufacturing technology; Siemens will provide engineers and expertise. Production should begin late next year.

Robert Pool

Red flag on CFC substitute

Washington

ONE of the most promising candidates to act as a replacement for ozone-destroying chlorofluorocarbons (CFCs) has been found to cause benign tumours in rats.

As a result, one US company has suspended marketing the chemical, HCFC-123, and a second has halted production of equipment designed to use it. Although Du Pont, the largest producer of CFCs, continues to use HCFC-123, the evidence does raise some 'red flags' and indicates that the process of introducing CFC replacements may not be as smooth as thought.

HCFC-123 had been hailed as a nearly ideal substitute for CFC-11, which is widely used as a coolant in 'chillers', large commercial and industrial airconditioning systems. The two substances have very similar chemical and physical properties, and although HCFC-123 is more expensive than CFC-11, it does only a fraction of the damage to the ozone layer.

As a final step in the testing of HCFC123, a consortium of CFC manufacturers paid for a two-year test of its effects on rats. Three groups of rats were exposed to varying levels of the chemical in the air they breathed, while a fourth, unexposed group acted as controls. The most highly exposed rats breathed air that contained 5,000 parts per million (p.p.m.) HCFC123 for six hours a day, five days a week, for two years; roughly equal to a human exposure of 10 p.p.m. for eight hours a day, five days a week, for 30 to 40 years.

At the end of the two years, 66 male rats in the highest-exposure group had developed 10 benign pancreatic tumours compared with none among 67 controls, and 14 benign tumours in the testes compared with four in the controls.

Except for the tumours, the rats exposed to HCFC-123 were actually healthier than the controls, says Anthony Vogelsberg at Du Pont. Among the control group, only 26 per cent were alive after two years, whereas 59 per cent of those exposed to 5,000 p.p.m. of HCFC-123 survived the study.

The apparent beneficial effects notwithstanding, the presence of the benign tumours signals a potential hazard of HCFC-123 use, and in a letter to the Environmental Protection Agency, scientists from the consortium of CFC producers said they were lowering their recommended maximum human exposure from 100 p.p.m. to 10 p.p.m.

"We see no problem with its use in industrial chillers or in any applications where the exposure can be kept below the [10 p.p.m.] level," Vogelsberg says. Du Pont has been using HCFC-123 to air-condition its own buildings, he says, and experience has shown that even in the area around the chillers, the levels of HCFC-123 are only 1 to 3 p.p.m.

Robert Pool

Selling research to save it

Hanover

ONLY a year ago, Fredrik Scheller and his group of molecular biologists at the East Berlin Academy of Sciences were content in the knowledge that their biosensor research was as good as anywhere in the world.

But in the newly reunified Germany, Scheller's group — like some 90 per cent of the researchers at the soon-to-be-disbanded academy — is facing imminent unemployment. Knowing that their research is world class is no longer enough. Now they must prove it.

Like a growing number of scientists in what was once East Germany, Scheller is preparing to put together a small company to sell the products of his group's research — and perhaps save some jobs in the process. The group's work on glucose-detecting enzyme membranes is both relatively practical and close to the cutting edge; properly packaged, priced and marketed, it might find buyers in the West.

Come October, ten such groups will test that premise. They are taking the plunge into the hard world of capitalist economics by participating in their first product exposition: the international biotechnology fair *Biotechnica 91*, to be held in Hanover. At the fair, they will try to sell that small proportion of East German research that is advanced enough and close enough to commercialization to find buyers on a world market.

For most of them, this is virgin territory. The few whose research had commercial application in the year before reunification last year had never had to worry much about world competition; the Soviet Union and the rest of the Eastern Bloc were essentially a captive market. And for the vast majority of the other state-sponsored researchers in East Berlin and the rest of East Germany, commercial potential of research was usually a distant concern; their jobs, after all, did not depend on coming up with marketable products. Anyway, without much modern instrumentation and equipment, most of the work was by necessity basic or theoretical — hardly the stuff to spin off blockbuster inventions.

Many of the eastern German researchers who are lucky enough to have some technology that might be marketed have applied to the new *Treuhandanstalt* agency for help with start-up money for a new company. Set up to sell off the assets of the former eastern state and to help eastern Germans start their own businesses with foreign capital, the *Treuhand* has the difficult job of deciding what technologies and products might possibly be successful in a Western market.

But the agency has found the going even

tougher than expected, as Western investors remain openly sceptical about the marketability of Eastern products. So far it has been able to find buyers for only about five per cent of the old East German companies — the cream of the crop — and takers are expected to become even rarer as the pickings get slimmer in the coming year.

Scheller, however, is fortunate enough to be able to bypass the *Treuhand* morass. He has found some partners with private capital and is hoping to rent a small laboratory in what, after December this year, will no longer be the academy but will be a smaller research institute that is expected to be called something such as the Research Centre for Biological and Medical Sciences.

There the group hopes to house a few technicians working on producing the enzyme membranes. Not a real research laboratory by any means, but enough to keep producing the materials and enzymes they already know how to make, and maybe, someday, start re-searching some new ones.

"It's not the greatest alternative," says team member Ulla Wohlenberger, "but it's all we have right now. It will be a very little company that will have no place for basic biosensor research in the early years."

Another new company entering the competitive world is the Berlin-based *FZB Biotechnik GmbH*, a year-old East Berlin company that specializes in optimized fermentation processes, enzymes and microbial cultures. Converted from a 500-person former government research centre to a much smaller private company last year, *FZB Biotechnik* is backed by *Treuhand* funding and the technical expertise of 170 biotech specialists.

Six other new eastern German biotechnology companies will also make their debut at the *Biotechnica* exposition, displaying their wares together in a joint stand with commercial arms of some 12 eastern German universities and research institutes. This will be their first opportunity of see how their fledgling marketing skills measure up to the established biotechnology powerhouses of the West. It may, as such, be a sobering experience.

"The main problem is that before, they only had to consider the technical part of their research, not the economic," says Karl Schügerl, a biotechnology expert at the University of Hanover.

"They had no problem selling what they had to the East in the controlled market. Now, of course, it's all different." Come October they will find out just how different that is.

Christopher Anderson

French scandal

London

Two former officials of the Institut de France are in jail awaiting trial for corruption as directors of the powerful research and analysis body prepare to reorganize its administration to avoid further such scandals. Stemming from an investigation in the spring by the *Cours des Comptes* government auditing agency, a total of four current and former institute employees have so far been charged with offences ranging from embezzlement to receiving stolen goods.

One of the officials, former technical adviser *Frédéric Gérard*, allegedly co-owned a holding company for a security company that was given an institute contract without proper public competition. The *Cours des Comptes* investigation also found "incredible" mismanagement at the institute, which includes the *Académie Française* and the academy of science, prompting promises of a complete management overhaul. C.A.

Australian brain drain

London

Australian scientists and academics are leaving the country at twice the rate of four years ago, according to a new survey by the Australian Bureau of Statistics. Last year the country lost 200 scientists, 50 academics and teachers, and 110 doctors, part of an exodus of more than 20,000 professionals and managers. Although Australia has set up a preferred-immigrant programme to attract foreign researchers and other professionals, it provides native academics and scientists with no special incentives to stay in the throes of the country's deepening recession. C.A.

RU-486 in UK

London

Known in the United Kingdom as *Mifegyne* and tightly controlled to prevent a black market, the abortion drug *RU-486* or whatever this is received its British licence last week and will soon be available on prescription.

Roussel-Uclaf, the drug's French manufacturer, has said that it will supply *RU-486* only after it has received a written request from an established hospital or clinic, and every pack will be coded so that its purchaser and prescriber can be tracked. It will also require that all doctors attend a special seminar before they will be allowed to prescribe it.

Private UK abortion clinics said last week that they would offer *RU-486* as an abortion option at £219 the same price as a surgical abortion. Three visits will be required: a counselling session, a visit to administer the drug and a final session at which the patient receives the drug *prostaglandin*, which induces contractions. C.A.

Personality put to the test

SIR — To paraphrase Samuel Johnson, it is no crime to write a foolish article. Accordingly, Blinkhorn and Johnson¹ walk the streets free men. But earlier correspondence² has not addressed the major and fallacious arguments of this paper.

We challenge Blinkhorn and Johnson's conclusion that there is "precious little evidence" that even the best personality tests predict job performance. We contend that (1) the three personality tests reviewed by these authors are far from the best personality measures in use; (2) the authors totally ignored important new developments in the field and presented simple and well-known statistical procedures as if they were revelations to researchers in personnel selection; (3) their highly contrived empirical illustration is so unrealistic as to be ludicrous; and (4) their informal survey of research was biased and inconsistent with other published surveys.

The three tests that Blinkhorn and Johnson singled out as representing the "top end of the market" are, on average, more than half a century old. A perusal of the most recent *Mental Measurements Yearbook*³ reveals harsh criticism of these tests, clearly demonstrating their inadequacy when evaluated by experts. There have been substantial developments in personality measurement that Blinkhorn and Johnson failed to reflect in their review. They contrived an example using 30 predictors with only 50 people, implying that this absurd attribute to entity ratio represented typical practice in personnel psychology. Such obvious misuse of statistics, even if isolated examples perpetrated by poorly trained or unscrupulous personnel consultants can be found, has no bearing whatsoever on the scientific question of whether or not personality measures predict job performance.

Blinkhorn and Johnson's conclusions were based by their own admission on "an informal survey of research". More systematic and comprehensive reviews⁴ using modern, quantitative meta-analytic methods have drawn more sanguine conclusions about the role of personality predictors of job performance. To cite but one research example, Day and Silverman⁵ demonstrated that even after considering the effects of cognitive ability, three personality dimensions were significantly linked to hypothesized facets of accounting performance.

Blinkhorn and Johnson failed to address the issue of utility. The appropriate evaluation of the utility of a selection system is not the level of the validity coefficient of individual predictors, but rather of the validity and benefit accruing from the use of composite predictors. Even with moderate validities, very substantial economic benefits have been demonstrated.⁶

Thus, it is obvious that Blinkhorn and Johnson began with the conclusion that per-

sonality testing had "insignificant" utility for employment selection and proceeded to "prove" their point by claiming that a small and unrepresentative set of studies and an example of an indefensible statistical procedure represented typical practice. On the contrary, systematically sampled validity studies from refereed journals require different conclusions: (i) when appropriate job analyses are undertaken, modern psychometrically sound personality measures are valid predictors of job performance, and (ii) such valid measures show substantial economic benefits when measured in terms of increased productivity and reduced training costs.

DOUGLAS N. JACKSON

Department of Psychology,
University of Western Ontario,
London, Canada N6A 5C2

MITCHELL ROTHSTEIN

School of Business Administration,
University of Western Ontario,
London, Canada N6A 3K7

1. Blinkhorn, S. & Johnson, C. *Nature* **348**, 671–672 (1990).
2. Gray, J. A. *Nature* **349**, 469 (1991).
3. Conoley, J. C. & Kramer, J. J. (eds) *The Tenth Mental Measurements Yearbook* (Buros Institute of Mental Measurements of the University of Nebraska, Lincoln, 1989).
4. Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. *Personn. Psych.* **37**, 407–422 (1984).
5. Day, D. V. & Silverman, S. B. *Personn. Psych.* **42**, 25–36 (1989).
6. Hunter, J. E. & Hunter, R. F. *Psychol. Bull.* **96**, 72–98 (1984).

Beating the systematics

SIR — Systematists are a vital part of the biological community, but conflicts continuously arise between the desire of systematists to apply the oldest names and the need of other biologists for nomenclatural stability (*Nature* **350**, 466; 1991).

We advance the following modest proposal with an eye towards making the interests of the two groups more convergent. Each time a change in nomenclature is published, the systematist should be required to conduct a literature search for all citations to the taxon in the past 50 years, and append the list to the paper.

This scholarly task would usually be child's play for those accustomed to ferreting out antecedent names from centuries-old, foreign-language documents. Furthermore, it is analogous to the requirement that the rest of us deposit vouchers of our taxa. But in rare cases the literature on a taxon would be so voluminous that the systematist would drop the task in favour of easier prey.

The names that would be saved in this way are precisely those that other biologists most want left alone, contributing stability to well-worked taxa while leaving the systematists free to exercise their art on taxa not yet discovered by the rest of us.

Of course there is no guarantee that a particularly zealous (or right-thinking, depending on one's point of view) name-changer would not go ahead and compile the complete list of all citations to, say, *Drosophila melanogaster*. But at least the rest of us could derive some solace from a useful appendix and also from the knowledge that its author must have gained some appreciation of the costs of such changes.

JOAN E. STRASSMANN

DAVID C. QUELLER

Department of Ecology and
Evolutionary Biology,
Rice University,
PO Box 1892,
Houston, Texas 77251, USA

Extremes of reality

SIR — I see that the war between the Flute Toodlers and the Tuba Boompers plays on, with the former insisting that, inasmuch as they cannot understand Wittgenstein, this proves he was crazy; the latter's indignant riposte is "What's wrong with being crazy?", as illustrated by F. A. Jenner (*Nature* **351**, 10; 1991), who maintains, with Karl Jaspers, that psychotics "see into depths" of reality that the merely sane cannot, offering us "an abundance of content representing fundamental problems of philosophy".

I would urge both parties to reconsider their extremist positions. The fact that we may not fully understand someone does not, necessarily, demonstrate that he is mad. Neither does it demonstrate that he is, necessarily, a profound and brilliant philosopher.

What entertains me are the arrogant absolutes indulged in on both sides, but what troubles me is the claim that, in dealing with questions of reality, insanity offers just as good and possibly an even better approach than sanity, these things merely being relative, as we all know.

No doubt, as Jaspers contentedly affirmed, "The philosopher in us cannot but be fascinated by this extraordinary reality [of the psychotic mind] and feel its challenge". Many millions felt its challenge in the 1930s and 1940s as a man gifted with such 'extraordinary reality' dealt with matters of reality in his own inimitable philosophic fashion.

Not understanding much of what Wittgenstein had to say leaves me in the position of being unable to decide definitively whether he was brilliant, loony or just possibly neither. But if I understand Dr Jenner of the University of Sheffield's Department of Psychiatry, I tremble for psychiatry's grip on reality.

RALPH ESTLING

The Old Parsonage,
Dowlish Wake,
Ilminster,
Somerset TA19 0NY

More views on Imanishi-Kari

Professor Mark Ptashne describes evidence to explain the disputed *Cell* paper, Dr Herman N Eisen (who conducted the MIT inquiry in 1986) concurs, Dr John Cairns (in letter to an unidentified official of the US National Academy) says the affair is the equivalent of Watergate and a former co-worker offers a testimonial to Imanishi-Kari.

From Mark Ptashne (Harvard)

In his account in *Nature*¹ of the MIT inquiry into the "Baltimore affair", Dr Herman Eisen raises one scientific point, and that point deserves some comment.

The central claim of Weaver *et al.*² is as follows: introduction into a mouse of a rearranged gene encoding an immunoglobulin heavy chain with a specific idiotype elicits production of antibody heavy chains that bear the transgene idiotype; remarkably, many of these chains are encoded entirely by endogenous genes. Why might one find such a result important? In the authors' words, the result "...suggests that a rearranged gene introduced into the germ line can activate powerful cellular regulatory influences." According to one interpretation, the result could be a manifestation of the workings of an "idiotype network"; the idiotype network idea, suggested several years ago, has not received strong experimental support.

Two papers published in 1989^{3,4} address the question of whether the central claim might be generally correct. Both of these papers report the use of a mouse to which has been added a rearranged heavy-chain immunoglobulin gene — the idiotype encoded by this gene is different from the one studied by Weaver *et al.*² As in Weaver *et al.*, however, the transgene in these new experiments encodes a heavy chain of type μ^a , whereas the endogenous μ gene is of type μ^b . The paper of Rath *et al.*⁴ shows that these transgenic mice do express antibodies with idiotype of the transgene (called AD8-reactive), but "...all detectable AD8 reactivity was associated with molecules expressing the μ^a allotype and none was detected in association with molecules lacking μ^a ." (Note, for example, that the idiotype was found neither on endogenous γ nor on endogenous μ chains.) Moreover, these idiotype-bearing μ^a chains were found in association with μ^b heavy chains in chimaeric molecules, an additional finding relevant to the controversy over Weaver *et al.*² as I explain below.

The paper of Durdick *et al.*³ finds that following immunization of these transgenic mice with the antigen that interacts with the

transgene product, idiotype can be found associated with γ heavy chains. Molecular analysis reveals, however, that the heavy chains of these antibodies are encoded by recombinants formed between the variable region of the transgene and an endogenous γ gene. These two papers thus fail to replicate, in an experiment involving a different idiotype, the central claim of Weaver *et al.*².

Dr Eisen, in his statement I alluded to at the beginning of this letter, says of Dr M. O'Toole, "But I did not agree with several of her arguments. For example, her assertion that some of the results could be explained by heterodimer formation, which I take to mean μ - γ chimaeric molecules is highly implausible". But O'Toole was referring not to μ - γ chimaeras but to μ - μ chimaeras: her memo to Eisen dated 6 June 1986, states "In sum-

mary, the data on transgenic sera in figures 1 and 2 can be explained...by heterodimer formation." In fact, Fig. 1 of Weaver *et al.*² involves testing only for μ chains, and Fig. 2 does not involve any isotyping. O'Toole's explanation, according to Rath *et al.*⁴, is eminently reasonable and readily testable.

I thank Herman Eisen for graciously reviewing this matter with me (see below). □

From Herman N. Eisen (MIT)

SIR — In my response¹ to Dr O'Toole's previous comments² I referred to her earlier suggestion (in a June 6, 1986 memo) that "heterodimer" formation could account for some of the reactivity in sera from the transgenic mice studied in the disputed paper in *Cell*³. I took the term heterodimer to mean a mixed μ - γ (or μ - α) dimer involving one μ ▶

To an officer of the National Academy of Sciences, 28 June 1991

AFTER our conversation, I thought I should produce a list of what I believe are the most important components of the so-called Baltimore affair.

(1) It seems that O'Toole was right in saying that the paper should be withdrawn (as eventually it was), right in thinking that there may have been misconduct (as the Secret Service now claim to have demonstrated) and right in her alternative interpretation for the central section of the paper (see the letter from Mark Ptashne).

(2) Nothing now is likely to stop the affair from progressing to its final disastrous conclusion, and I do not see how the authors of the paper can escape public censure, at the very least. About the only question remaining is whether anyone will actually go to jail.

(3) The whole affair seems to be turning into a kind of scientific Watergate and, like Watergate, is surely destined to be dissected and analysed for years to come.

(4) Some of the blame falls on the scientific community — on those who arranged and conducted the initial, perfunctory inquiries — on the National Academy for not demanding a proper investigation — and on the many scientists who did not look at the evidence and, instead, construed the whole business as a Congressional manoeuvre to attack the scientific establishment. (I remember that originally I too felt that the row was probably a political stunt.)

(5) Because the establishment has played such an undistinguished role, we may find it increasingly difficult to maintain the idea

that science is a genuine search for truth and that scientists are generally honourable and deserving members of society. Simply at the mundane level of money, I could imagine fund-raising for the Academy becoming much harder if Congress is left with the image of the Academy as the organization that sided with Baltimore right or wrong, through thick and thin, to the bitter end.

(6) So I believe that, although it now too late to do much good, the Academy should be issuing a statement (a) reaffirming the aims of science and (b) pointing out that if the rules and principles of science had been observed we wouldn't now be in this mess. For most scientists, science is the pursuit of a truth that is external to our wishes. This truth is quite unlike the verdict of a court of law because it does not depend on advocacy. Instead, each of us has to be responsible for the accuracy of our own statements; we cannot simply count on others to correct our mistakes. Each of us knows more about our own experiments than anyone else, and when something goes wrong we have to speak up. If the Academy does not say something like that, American scientists may end up with the same kind of public image as many of the country's lawyers and politicians — which would do a great disservice to all young scientists.

JOHN CAIRNS

Department of Cancer Biology,
Harvard School of Public Health,
665 Huntington Avenue,
Boston, Massachusetts 02115, USA

1. Eisen, H. *Nature* **351**, 343 (1991).

2. Weaver, D., Reis, M. H., Albarese, C., Costantini, F., Baltimore, D. & Imanishi Kari, F. *Cell* **45**, 247–259 (1986).

3. Durdick, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 2346–2350 (1989).

4. Rath, S., Durdick, J., Gerstein, R. M., Selsing, E. & Nisonoff, A. *J. Immun.* **113**, 2074–2080 (1989).

5. O'Toole, M. *Nature* **351**, 180–183 (1991).

chain encoded by the transgene and one γ (or α) chain encoded by an endogenous gene. Such a heterodimer seemed unlikely to be stable because of extensive amino-acid sequence differences between μ and other types of immunoglobulin heavy chains. Recently, however, it has been called to my attention (see Prashne's accompanying letter) that what O'Toole meant by heterodimer was not μ - γ (or μ - α) but a mixed IgM molecule (a polymer with 10 chains per molecule) in which some of the μ chains were encoded by the transgene and others by endogenous μ genes. In contrast to heterodimers, such mixed μ - μ heteropolymers are certainly plausible and O'Toole's suggestion would thus appear to have been reasonable.

Nonetheless, there was no guarantee at the time that the critical antigenic determinants of the transgene's heavy chain ("idiotype") would be manifest in mixed IgM molecules, where various μ chains can interact. Hence, even if there had been no misunderstanding about the matter in the 1986 inquiry, I would nevertheless have felt that further research was needed to evaluate her proposed explanation. Some of the further research has indeed been reported in a paper from Alfred Nisonoff's laboratory (Rath *et al.*⁴). The paper contains some elegant analytical immunochemistry showing that in another transgenic mouse strain, involving a different μ chain transgene, mixed IgM molecules (transgene μ chains plus endogenous gene μ in various proportions) do exist; it also shows that in these molecules the transgene's idiotype is manifest and is due exclusively to the transgene's μ chain. The results in Rath *et al.* may well portend what would be found if sera from the transgenic mice used in the disputed *Cell* paper (M54 and M95 strains) were similarly analysed.

There are, however, several reasons for exercising caution in extrapolating now from the study by Rath *et al.* to the *Cell* paper. (1) The *Cell* paper claimed that some hybridomas did not produce IgM (but presumably produced IgG or IgA) and were still idiotype positive. The existence of such hybridomas is still under dispute; for such hybridomas mixed μ - μ molecules would not be relevant. (2) In Rath *et al.* the transgene's idiotype was detected by a monoclonal antibody, whereas in the *Cell* paper it was detected by a polyclonal antibody population; the range of reactivities (or cross-reactivities) exhibited by polyclonal antibodies are expected to be substantially broader than those of any particular monoclonal antibody. (3) It is not yet known what endogenous immunoglobulin genes are expressed in the transgenic mouse strain studied by Rath *et al.* In the strains studied in the *Cell* paper these genes appear to be confined to an unusual, limited set at the extreme 3' end of the huge array of V_H gene segments (represented by the V81X family). The point here is that immunoglobulins whose variable (heavy chain) domains are encoded by V81X-like genes have been claimed by several laboratories to be highly

cross-reactive with diverse reagents (including many anti-idiotypes). Although the validity of this claim is still not settled, it is important to keep in mind that diverse transgenic mouse strains might express different sets of endogenous immunoglobulin genes and thus differ in their reactivities.

In the excellent paper by Durdik *et al.*³ hyperimmunized transgenic mice were found to express many recombinant genes (linking the variable sequence of the transgene to the constant sequence of an endogenous (γ) gene. Their immunization programme also resulted in strong selection of those B cells that expressed both the transgene and a particular light chain gene (required for detection of a particular form of the idiotype). The relevance of this paper to the *Cell* paper (which also described a hybridoma with a similar recombinant gene) is limited, because in the *Cell* paper only nonimmunized mice were analysed.

Besides being elegant studies in their own right, Rath *et al.* and Durdik *et al.* are welcome demonstrations that the scientific process itself is the most effective way of resolving scientific disputes. I am grateful to Professor Mark Ptashne for calling these papers to my attention and for suggesting that my comments accompany his. I look forward to additional research that bears on disputed scientific issues in the *Cell* paper. □

From Nicholas Yannoutsos (NIMR, Mill Hill)

I SHOULD like to comment on the investigation of the paper published in *Cell* in 1986 whose principal author was Dr Thereza Imanishi-Kari. I worked with Imanishi-Kari from about October 1985 to May 1988 and I feel it is my duty to put on record my personal experience of that period.

My work was mainly the establishment of transfectant cell lines and transgenic mouse lines carrying the membraneless form of the 17.2.25 immunoglobulin heavy-chain gene, their molecular, serological and FACS analysis, and comparison to cell lines that carried the intact gene versus non-transfected cell lines and normal mice. This was always compared with a parallel analysis of the transgenic mouse line which carried the intact gene and whose serological, pre-B and hybridoma study had been reported in the 1986 *Cell* paper.

My work, like most of the work in Imanishi-Kari's laboratory, was based on the findings reported in that paper and I want to make clear that, in my personal experience, what was reported in that paper was not an isolated collection of ambiguous experiments. Instead, it was part of continuing research conducted with genuine and critical interest. I must stress the "critical interest" because Imanishi-Kari herself and other people in her laboratory, including myself, have painstakingly repeated time and again the work reported in the *Cell* paper. This was done with improved and diverse techniques and approaches and alongside further

experiments that might clarify the mechanisms involved. All this was done in an atmosphere of openness and intellectual integrity that kept everybody alert to the possibility of trivial or artefactual explanations for what was obviously a profound effect on the immunology of the transgenic mouse under study. Not only did all the people in the laboratory participate actively, but so did people from collaborating or just neighbouring laboratories. The data were scrutinized, discussed and analysed extensively, in group and departmental meetings or even as they were coming out "raw" in the corridor and at the benches of MIT and Tufts.

I cannot emphasize enough the genuine attitude with which Imanishi-Kari conducted her own work and invited other people's participation, criticism and contribution, and her willingness to pursue not a particular theory, but any valid interpretation for her own findings and the findings of other researchers in related work. She was always in the laboratory, working long and hard hours and in constant communication with the people in it. She was particularly strict about the technical aspects of the work and demanded that every experiment be well controlled and repeatable. Her strictness might occasionally frustrate the false pride of an individual worker, but it was also a lesson in the essential modesty, dedication and correctness with which scientific work must be conducted and of which Dr Imanishi-Kari was herself the best example.

Several aspects of the originally reported work are repeated and further analysed in a recent publication (Iacomini *et al.*, *Int. Immun.* 3, 185-196; 1990). Among the experiments reported in this publication is the Abelson transformant pre-B analysis of the transgenic mouse carrying the membraneless form of the 27.2.25 Ig heavy chain gene. As I mentioned above, I have worked with this mouse and compared it to the original transgenic mouse. In the process of such comparison (unpublished), I have done RIAs in which sera from both transgenic mice were screened with anti- λ , anti- κ , anti- μ a (BET-1 and anti- μ b antibodies on plates coated with monoclonal or polyclonal anti-17.2.25 anti-idiotype antibody. So the disputed serology on the original transgenic mouse was repeated and shown to be essentially as reported in the 1986 *Cell* paper.

Since the beginning of this affair, the investigating committees and the scientific journals that have been reporting it appear to have focused only on the state of the notebooks that contain the initial experiments on this transgenic mouse. No due consideration seems to have been given to whether the reported findings are actually valid and independently reproducible. Despite the extensive coverage of this case, it is quite unclear to me, and no doubt to most people familiar with the work, what were the real scientific grounds for the retraction of the 1986 *Cell* paper by several (but not all) of its authors, or in fact for this whole "Baltimore affair". □

Character analysis of a science

Eric Magnusson

Chemists, like other specialists, ought to make it clearer what hypotheses they are framing and testing. They will thus avoid the suggestion that they are over-fond of empirical generalizations.

Do chemists deserve the reputation in some quarters that they are over-fond of elevating generalizations into natural laws? I have taken a sample of 167 papers from four successive issues of the *Journal of the American Chemical Society* published in 1990, together with the 73 general articles published in four issues of *Nature* in 1990. I assigned papers to four categories: high-level explanatory theories, empirical generalizations, observations and technological reports — this last category consisted of papers in which known chemical principles have been applied to technology (Table 1). I also ranked my sample for generality of subject matter, using an arbitrary scale and again including a category for technological products (Table 2). As I anticipated, the papers in both samples described highly specialized work: because their subject matter was restricted in range, so were their conclusions.

The objectives of the chemists writing in *JACS*, a refereed, general-interest journal, surprised me: almost half were trying to find out the mechanisms of chemical reactions. If this journal is any guide, chemistry is well past the stage where the most effort goes into making new compounds and determining their chemical structure. These tasks now consume only 18% (synthetic) and 14% (structural) of the sample. Papers about the internal structure of molecules (electronic structure) represented 11%, and the rest (14%) were mainly devoted to measuring and explaining physical properties, including spectroscopy and general chemistry.

Unless the clearly unrepresentative nature of the sample invalidates my conclusions, I judge that chemistry is a mature science — almost as many contributors to *JACS* are seeking high-level explanations as those in *Nature*. Although articles in *Nature* tend to report major discoveries and *JACS* papers continuing projects, there is a remarkable similarity in the explanatory process which the two sets of contributors follow.

Despite Maddox's impression¹, my analysis revealed very little daring in the *JACS* generalizations. Most of them were the reverse — explanations in the process of being made. Because the explanations still lacked sufficient corroboration, the claims had to be muted. Thus, several papers dealt with biological processes involving compounds too fragile to isolate. So the researchers made model molecules and reported on the similarities and regularities in the behaviour of the natural and artificial compounds². Ultimately the high-level explanation will come, but not yet.

It is true that chemists do not seem to move in the classical manner, from observation to generalization to explanation. The pattern seems to be that the explanation is known in advance and is adapted to the phenomenon in question. Most of the conclusions classified as 'generalizations' gained that category by default: they were qualitative; neither bare observations nor explanations.

Surprisingly, my sample contained no mathematically fitted generalizations. There were however, generalizations of the classical kind: genuinely predictive statements, like Kepler's laws, with important implications for the ultimate explanation but with no

TABLE 1 PAPERS RANKED BY LEVEL OF UNDERSTANDING

	<i>JACS</i>	<i>Nature</i>
Explanatory theory	44	51
Generalization	47	34
Observations	7	12
Technological report	2	3

Values in per cent.

TABLE 2 PAPERS RANKED BY GENERALITY OF SUBJECT MATTER

	<i>JACS</i>	<i>Nature</i>
Perfectly general	3	6
Restricted class	4	3
Single category	90	89
Technological product	3	3

Values in per cent.

real explanatory content themselves. For example, the chemists who measured the Langmuir adsorption isotherms of mixtures of optical isomers on a column³ found two kinds of sites, one with a preference for the D isomer, the other indifferent. The separation of the isomers is partially explained but the explanation is not chemical and does not reach my "explanatory theory" designation.

A chemical explanation is usually an explanation at the atomic, molecular or electronic level. An example is the explanation of the effectiveness of a technique to attach a substituent to a specific site in a steroid molecule by calculation of the 'stiffness' of the transition state in the reaction⁴. This is a 'molecular' explanation: classical 'molecular mechanics' is sufficient to obtain it. Chemists do indeed "want a picture of what has happened"¹. 'Chemical' explanations are often pictorial, or if quantum phenomena intervene to make that impossible, at least describable by means of standard concepts.

Explanations at the electronic level have become much more common in the chemical literature now that 'user-friendly' electronic structure packages are so readily available. Specialist investigations of electronic structure represented 11% of my *JACS* sample,

for example the use of geometry optimization (searching the conformational 'space' to find the minimum energy geometry for a molecule) to resolve a discrepancy between the theoretical and experimental structures reported for vinyl alcohol. The theoretical 'experiment' led to re-analysis of the experimental data, so removing the discrepancy⁵.

It is clear that chemists are not "elevating empirical generalizations into natural laws": their explanations are visions about molecular motions and electron clouds that have come down from the heavens. The explanatory theories of the gods (quantum mechanics and statistical mechanics) are too hard for the real chemical world, so chemists have made approximate models of them, separately adapted for each application. Hartree-Fock/Born-Oppenheimer electron-cloud theory is one; molecular mechanics is another; the collision theory of what makes molecules reactive is another. Chemistry, very much a theory-driven discipline, uses these models all the time.

If the *JACS* sample is a proper guide, the chemists who do the experiments are the ones who make the explanations. I did not find any papers by theorists making explanations for other people's data. High-level explanations were almost always supported by observational evidence in the same paper. In many cases, the empirical generalization stage was left out altogether.

Chemists should be grateful to Maddox for reminding them that their science has a structure. After wading through the papers in a sample this size, it occurs to me that chemistry would profit if chemists paid more attention to the character of their science, saying explicitly what hypothesis they have tested, what the tests were and how well it survived. Making science understandable to non-specialists is an important responsibility, but making it assessable is also important. Away from the physical sciences, the processing of 'soft' data forces researchers to pay stringent attention to validity and reliability. It may be time for the rest of us also to acknowledge that our science is provisional, and to learn the proper way to frame, test and report hypotheses. □

Eric Magnusson is in the Chemistry Department, University College, University of New South Wales, Campbell, ACT 2600, Australia.

- Maddox, J. *Nature* **347**, 119 (1990).
- Wasielowski, M. R. et al. *J. Am. chem. Soc.* **112**, 6482 (1990).
- Jacobson, S. et al. *J. Am. chem. Soc.* **112**, 6492 (1990).
- White, P. & Breslow, R. *J. Am. chem. Soc.* **112**, 6842 (1990).
- Smith, B. J. & Radom, J. *J. Am. chem. Soc.* **112**, 7525 (1990).

Déjà vu all over again

Gregory A. Petsko

ALTHOUGH new protein folding motifs keep being discovered¹, it is much more likely that any 'new' protein structure will, in whole or in large part, look very similar to a structure that has previously been determined. The first question any structural biologist asks on being told that a new structure has been solved is no longer "What does it look like?"; it is now "What does it look like?". And the most prevalent sensation in structural biology is that best described by the noted American baseball player and philosopher, Yogi Berra: "Déjà vu all over again". On pages 168 and 172 of this issue, Schiering *et al.*² and Kuriyan *et al.*³ report two new enzyme structures that enlarge the scope and importance of one protein structural family: that of the flavin-dependent disulphide oxidoreductases. These structures, together with two others recently determined^{4,5}, indicate that this family is capable of striking versatility in its chemical behaviour and also illuminate some details of the process of catalytic evolution.

Flavin-dependent disulphide oxidoreductases use the isoalloxazine ring of flavin adenine dinucleotide (FAD) to shuttle reducing equivalents from NADPH (or NADH, depending on the enzyme) to a cysteine residue that is usually part of a redox-active disulphide bridge. In a second step, the reduced disulphide reduces the substrate. This is a simple theme, but the fun is in the variations.

Six members of the family have had their structures determined by X-ray diffraction, and they cover an impressive range of catalytic — and biological — functions. Glutathione reductase ensures that the cell has enough reduced glutathione to maintain protein thiol groups in the reduced state⁶. Trypanothione reductase carries out the analogous reaction in trypanosomal cells; its substrate, trypanothione, is simply oxidized glutathione with a spermidine crosslink between the two carboxyl groups of the glycines of glutathione⁷. Lipoamide dehydrogenase, part of the 2-oxo acid dehydrogenase multi-enzyme complex, oxidizes the dihydrolipoil groups of lipoate acyltransferase and so couples glycolysis to the tricarboxylic acid cycle⁸. Thioredoxin reductase catalyses the reduction of a disulphide bond in a protein, thioredoxin, which has a part in the redox metabolism of plant and animal cells, including participation in DNA synthesis⁹. Mercuric reductase enables bacteria to detoxify the mercuric ion by reducing it to elemental mercury, which evaporates from the cell¹⁰. And NADH peroxidase carries out the two-electron reduction of hydrogen peroxide to water¹¹.

Glutathione reductase is the structural parent of the family. It is dimeric, having four domains per monomer: an N-terminal

FAD-binding domain, an NADPH-binding domain, a central domain, and a C-terminal interface domain (see figure). Glutathione reductase has a preference for NADPH over NADH, but, with the structural data^{12,13} as a guide, Scutcheon *et al.* reversed this coenzyme specificity by means of seven point mutations¹⁴. Substrate binds in a deep crevice at the junction of the FAD-binding and central

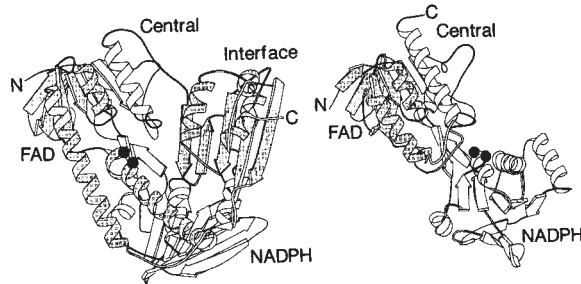


Diagram of tertiary structures of human glutathione reductase¹³ (left) and *Escherichia coli* thioredoxin reductase³ (right). Glutathione reductase has four structural domains: the FAD- and NADPH-binding domains, the central domain and the interface domain⁶. The FAD and interface domains in glutathione reductase are shaded, as is the FAD domain in thioredoxin reductase, which lacks the interface domain³. Both enzymes are dimeric in the active form; only the monomers are shown here. The FAD and interface domains of both enzymes are shown in the same orientation. The NADPH domains have similar chain folds, but are displaced by a relative rotation of 66° (ref. 3). The redox-active cysteines are indicated by solid circles. The chain fold of mercuric reductase is similar to that of glutathione reductase, and is shown in this orientation in Fig. 2 of ref. 2. (Figure prepared by T. S. R. Krishna and J. Kuriyan, using a program by J. Priestle.)

domains of one monomer plus the interface domain of the other monomer¹². (Enzyme active sites are usually found at interdomain and/or intersubunit regions.)

Other members of the family show a variety of structural differences which provide some intriguing clues as to how new enzymes evolve from older ones. Most of the four glutathione reductase domains appear in the other enzymes in the family, but their relative orientations are often different. Since the active site occurs between the domains, this alteration provides a simple mechanism for reconfiguring the catalytic and specificity properties to meet the special requirements of each reaction. For example, lipoamide dehydrogenase has an extension at the C terminus relative to glutathione reductase¹⁵. This extra stretch of chain makes the binding site smaller to accommodate the smaller substrate lipoate. Relative to glutathione reductase, the domains are rotated by about 7°, opening up the dimer interface by about 4 Å (ref. 16).

Examination of the structures of other members of the family shows that this flexibility of domain positioning is characteristic. NADH peroxidase and trypanothione reductase also have four domains that are

individually similar to those of glutathione reductase, but their relative orientations are different. In NADH peroxidase the rearrangement is a concerted rotation of the FAD-binding, NADH-binding and central domains by 18° relative to the fixed position of the interface domains⁴, leading to a more open domain arrangement. An open active site may be important for ease of release of the product water molecules. Further, unlike all other members of the family, the redox-active thiol of NADH peroxidase is not involved in a disulphide bridge. Instead, the reactive cysteine is in the form of a stable cysteine-sulphenic acid¹⁷. Presumably, the need to stabilize this normally short-lived species is reflected somehow in the rearranged active-site region.

In trypanothione reductase there is a similar, but smaller, rotation by 4° of the FAD, NADPH and central domains away from the interface domain of the other monomer⁵. The result is the formation of a substrate-binding site that is very similar to that of glutathione reductase at the base (near the FAD and the redox-active disulphide) but 4 Å wider in the region where the spermidine crosslink of trypanothione binds. Additional, local amino-acid changes convert this upper region from a positively charged site in glutathione reductase to a more neutral one. Reduction in charge would seem to be needed here, as trypanothione

lacks the free carboxylates of glutathione. The structural similarity of glutathione reductase and trypanothione reductase has allowed Bradley *et al.*¹⁸ and Henderson *et al.*¹⁹ to switch the substrate specificity of the first from glutathione to trypanothione by changing the polarity of this upper pocket.

The most dramatic evolutionary changes are found in the structures of mercuric reductase² and thioredoxin reductase³. In the first of these, 166 amino acids are added onto the N terminus of the glutathione reductase framework. The extra segment is actually a tandem repeat of an 80-residue domain that is highly homologous to the periplasmic mercury-binding protein Mer P (ref. 20). The two domains can be cleaved from mercuric reductase without altering its enzymatic properties, and perhaps it is not surprising, therefore, that the extra 166 residues are not clearly represented in the electron density map because of their high mobility. The core structure of mercuric reductase is nearly the same as that of glutathione reductase, but 15 residues, containing a second redox-active disulphide, are added to the C-terminal interface region. Because the active site is at the subunit interface, the primary redox-active disulphide of one subunit is close to

the additional disulphide at the C terminus of the adjacent subunit. Mercury ligation by two cysteines from different disulphides could lead to longer S-Hg bonds and may, in combination with the unusual additional ligation by two tyrosines, help prevent formation of a dead-end S-Hg-S complex.

In all of these flavoprotein oxidoreductases, the polypeptide-chain folds of the NADPH (or NADH)-binding domains are similar to those of the FAD-binding domains, suggesting that these domains evolved by gene duplication²¹. Moreover, the structural data are consistent with divergent evolution from a common ancestor (perhaps lipoamide dehydrogenase since it is used in anaerobic metabolism).

The structure of thioredoxin reductase, however, shows that the family could have evolved by more sophisticated mechanisms. Although the individual structures for the FAD, NADPH and central domains are the same, thioredoxin reductase lacks the interface domain completely. As a result, the NADPH-binding domain is rotated by 66° relative to its position with respect to the FAD-binding domain in glutathione reductase (see figure). Thioredoxin, the protein substrate, binds to a shallow depression at the dimer interface, where the redox-active disulphide is located. This disulphide is on the opposite side of the flavin to the disulphides of the other enzymes in the family, and was probably acquired independently. A reasonable model for the evolution of this enzyme would thus involve both convergent and divergent evolution operating on the same gene. Thioredoxin reductase may have been assembled in its present form by accretion of structure modules, or it may have resulted from the deletion of the interface domain in a regular family member. Whichever occurred, the resulting molecule would not have had its redox-active thiol in the

correct location for either, say, lipoamide dehydrogenation or thioredoxin reduction. That component of the active site was either already present by accident, or was acquired later, resulting in convergence to the familiar chemical function.

Modular protein construction provides a rationalization for the prevalence of interfacial active sites. The relative positions of the catalytic and substrate-binding residues can be altered by random mutations far from the active site if such mutations change the intersubunit or interdomain contacts of modules. Especially when more than one substrate or coenzyme is involved, this process allows considerable alteration in the rel-

ative positions of active-site residues without risk of loss of binding, because that is provided by the domains themselves.

Although the flavin-dependent disulphide oxidoreductase family has only six members so far, it has proven to be an abundant source of information. And even the same old motif, it seems, can be used in surprising ways. More surprises will no doubt follow as new structures enlarge the family. Yogi had a saying about that, too: "It ain't over 'til it's over". □

Gregory A. Petsko is in the Structural Biology Laboratory, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254, USA.

MANTLE STRUCTURE

A matter for resolution

Thorne Lay

THE most conspicuous features of the Earth's surface are the continents and the ocean basins which separate them. According to Su and Dziewonski, on page 121 of this issue¹, the Earth's mantle, extending from a depth of a few tens of kilometres to 2,900 km, is dominated by similarly large features. Pro-

only be partially deduced from its surface expression. The most direct procedure for constraining the mantle's three-dimensional structure is seismic tomography, in which material properties of the interior are inferred from elastic-wave velocity heterogeneity. Seismic-wave travel times are used

to image the heterogeneity, with depth resolution being obtained by analysis of various wave types that sample different portions of the interior.

Su and Dziewonski constrain the heterogeneity scales in the mantle using the seismic phase SS, a shear wave that is reflected once off the Earth's surface on the path from the seismic source to the receiver (Fig. 1). The relatively uniform global coverage provided by this phase, and its sensitivity to the velocity structure in the upper mantle under the mid-path reflection point, allow the construction of shear-velocity heterogeneity spectrum expanded up to spherical harmonic degree 36 (scale lengths larger than 1,000 km). The authors find that the spectrum is dominated by large-scale components (degrees 1 to 6) with corresponding spatial dimensions greater than 6,000

km. Under the common assumption that lateral shear velocity variations are thermally induced, and hence related to convective motions, Su and Dziewonski argue that mantle convection is domesticated by correspondingly large-scale components.

A similar picture of large-scale features in the upper mantle comes from other recent seismic studies^{2,3}. The heterogeneity spectrum² (Fig. 2) determined using 100-second-period Love waves, surface waves that keep to the Earth's top 200 kilometres, resembles



FIG. 1 Polar cross-section through the tomographic picture of the mantle deduced by Su and Dziewonski from shear-wave arrival times. Note the left-right asymmetry in the upper mantle, caused by the presence of continents on the right side of the cross-section, and their absence on the left. (Grey scale spans a ± 3 and ± 1 per cent variation in velocity in the upper and lower mantle, respectively.)

duced by thermal contrasts that are associated with the mantle's convection, these structures may be clues to the behaviour of the mantle. But on the other hand, are we overlooking the smaller details that may actually be more significant?

The existence of large tectonic plates and of concentrated flow features such as subducting slabs and ascending plumes beneath hotspots require there to be both large- and small-scale components of mantle flow, but the internal convective configuration can

1. Stallings, W. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5046-5050 (1991).
2. Schiering, N. et al. *Nature* **352**, 168-172 (1991).
3. Kuriyan, J. et al. *Nature* **352**, 172-174 (1991).
4. Stehle, T., Ahmed, S. A., Claiborne, A. & Schulz, G. E. *J. molec. Biol.* (in the press).
5. Kuriyan, J. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
6. Schirmer, R. H. & Schulz, G. E. in *Coenzymes and Cofactors* Vol. 2 (eds Dolphin, D., Poulson, R. & Avramovic, O.) part B, 333-379 (Wiley, New York, 1987).
7. Shames, S. L., Fairlamb, A. H., Cerami, A. & Walsh, C. T. *Biochemistry* **25**, 3519-3526 (1986).
8. Perham, R. N., Harrison, R. A. & Brown, J. P. *Biochem. Soc. Trans.* **6**, 47-50 (1978).
9. Holmgren, A. *J. biol. Chem.* **264**, 13963-13966 (1989).
10. Fox, B. & Walsh, C. T. *J. biol. Chem.* **257**, 2498-2503 (1982).
11. Dolin, M. I. *J. biol. Chem.* **250**, 310-317 (1975).
12. Pai, E. F. & Schulz, G. E. *J. molec. Biol.* **258**, 1751-1757 (1983).
13. Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **195**, 701-729 (1987).
14. Scrutton, N. S., Berry, A. & Perham, R. N. *Nature* **343**, 38-43 (1990).
15. Schierbeek, A. J. et al. *J. molec. Biol.* **206**, 365-379 (1989).
16. Mattevi, A., Schierbeek, A. J. & Hol, W. G. J. *J. molec. Biol.* (in the press).
17. Poole, L. B. & Claiborne, A. *J. biol. Chem.* **264**, 12330-12338 (1989).
18. Bradley, M., Büchele, V. S. & Walsh, C. T. *Biochemistry* **30**, 6124-6127 (1991).
19. Henderson, G. B. et al. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
20. Wang, Y. et al. *J. Bacteriol.* **171**, 83-92 (1989).
21. Schulz, G. E. *J. molec. Biol.* **138**, 335-347 (1980).

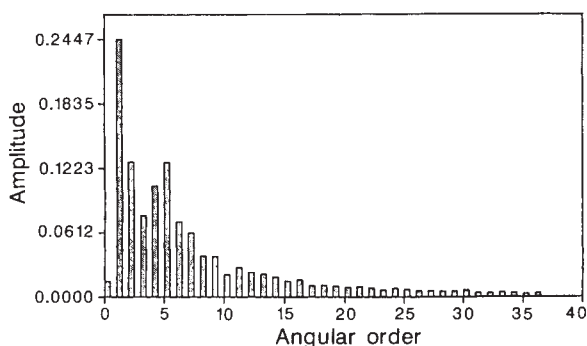


FIG. 2 Spectral amplitude of 100-second-period Love (surface) wave phase-velocity variations. (From ref. 2.)

Su and Dziewonski's SS spectrum in many ways. The strong degree-1 term results from the gross distribution of mean crustal thickness caused by the preponderance of continents in the Northern Hemisphere (Fig. 1); the peak at degree 5 is a chance result of the particular distribution of shields, old oceans, hotspots and ridges³.

The apparent dominance of these large-scale structures is surprising given that the strongest temperature and velocity contrasts occur at much smaller scales. For example, subducting, cold lithosphere produces the strongest lateral contrasts in the mantle, with temperature drops of 1,000 K possible over less than 100 km. Although accompanied by a 3–7 per cent change in seismic velocity, such contrasts are poorly manifested in the SS tomography¹.

Thermal plumes, hot structures that rise through the mantle to produce volcanic hotspots such as Hawaii, have 100 km scale lengths in their deep conduits and 500 to 1,000 km scales in their flattened plume heads. The number of volcanic hotspots at the Earth's surface might suggest that these plumes should generate a dominant scale of mantle heterogeneity.

Lastly, regions of partial melting below back arc basins and oceanic ridges can produce extreme variations in seismic shear velocity, especially near magma chambers. Again these may be expected to produce velocity variations over scales of several hundred kilometres, but these are simply not resolved by Su and Dziewonski's technique.

Instead, it is the large-scale structures, exceeding 1,000 km in dimension, that are resolved in the current generation of seismic models. The dominance of 6,000-km and larger structures appears to reflect smooth global variations such as the progressively thickening thermal boundary layers in oceanic lithosphere, and the transitions within the continents from rigid craton to mobile belt. The lesser importance of smaller structures has fundamental implications for the complexity of thermal structure in the convecting model that is yet to be fully quantified. It seems the system is not chaotic with uniform heterogeneity at all scales.

But both the SS and Love-wave heterogeneity spectra are strongly influenced by shallow mantle variations in the upper 200 km, within the near-surface thermal and

chemical boundary layer. It is well known that the strong lateral thermal gradients within a boundary layer produce enhanced power in seismically detectable heterogeneity⁴, and thus these results alone cannot be used to infer the deeper mantle dynamic configuration.

Consideration of this third dimension further complicates the question of the dominant scale in mantle heterogeneity.

As can be seen in Fig. 3a, a much greater range of seismic velocities is found in the upper mantle than in the lower mantle, especially for S waves⁵. The SS experiment probably cannot see through these strong shallow variations to resolve the structure of weaker, deep heterogeneity. Except at the upper and lower boundaries, there is little variation in seismic velocity in the lower mantle (Fig. 3b). Tomographic models^{5–7} do suggest, however, that the variations occur mostly at the large length scale, with the spectrum being peaked at degrees 1 and 2, and with little contribution above degree 6. This suggests that large-scale flow features also dominate in the lower mantle, producing the components⁸ in the geoid (the gravitational map of the world) with wavelengths greater than 10,000 km.

Given the preponderance of evidence for large-scale lateral structure in both the upper and lower mantles what can be said about the radial structure of mantle dynamics? It is generally agreed that the entire mantle is a thermal convecting system, with about 10 per cent of the heat coming from the core and 90 per cent due to internal radioactive decay. What is unresolved is whether it is radially layered owing to compositional stratification, and whether the surface plates are driven by large-scale circulation or by distributed plume flow (see also the discussion by Garwin⁹). Whole-mantle flow intrinsically implies large-scale structures, although small-scale thermal plumes may contribute to the circulation, but the velocity heterogeneity spectra alone do not resolve the question of layered- versus whole-mantle convection.

Even if whole-mantle convection occurs, a difference in the dynamics of the shallow and deep mantles is expected given the strong evidence of a significant increase in viscosity near the 670 km seismic discontinuity¹⁰. Whole-mantle convection models with greater chemical heterogeneity in the deep mantle that endures because of the slow convective stirring of the high-viscosity region¹¹, are not consistent with the low level of heterogeneity throughout the region unless the chemical variations are seismically

invisible. It is interesting to note³ that the strong degree-2 patterns of the shallow- and deep-mantle heterogeneity are spatially shifted across a depth of 800 km. The material below 800 km has substantially reduced velocity heterogeneity, but the relative degree of thermal and chemical heterogeneity is unclear because of uncertainties in the high-pressure variation of seismic velocity with temperature and composition. Evidence of thickening and deflection of subducting slabs above 800 km depth is accumulating, thus the possibility that the mantle has a stratified convective system with a chemical boundary deeper than the phase transition near 670 km must still be considered. Relatively large aspect ratio convection is implied if the upper 800 km is a separate layer, but the influence of the high-viscosity surface plates could allow stable layered mantle configurations with a long-wavelength heterogeneity spectrum¹².

From the seismological perspective, the predominance of the long wavelengths is

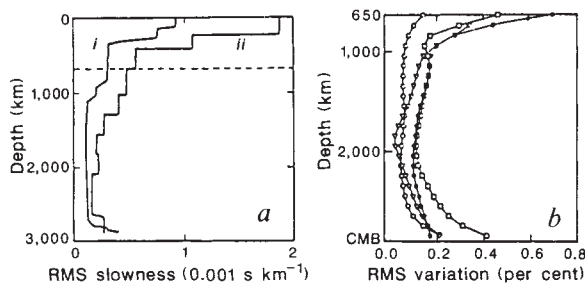


FIG. 3 a, Root-mean-square variations for the slowness of seismic waves for the whole mantle: i, P waves⁵ and ii, S waves⁶. The strong shallow heterogeneity in the velocity of P-waves seems to extend over a wide range of length scales. b, Percentage root-mean-square slowness variations for lower-mantle velocity models (▽, from ref. 5; □, ref. 6; ●, ref. 7; ○, ref. 8). CMB, core-mantle boundary.

helpful, as it permits the use of many of the asymptotic procedures of seismic tomography. But, a complete characterization of mantle heterogeneity extending to the very short scale-lengths of dynamically significant slabs and plumes is far from accomplished. Until the power in the high-degree terms is accurately resolved, any inferences about the dominant scales of mantle dynamics must be viewed as tentative. □

Thorne Lay is in the Earth Science Department, University of California, Santa Cruz, California 95064, USA.

1. Su, W. & Dziewonski, A. M. *Nature* **352**, 121–126 (1991).
2. Zhang, Y.-S. & Tanimoto, T. *Phys. Earth planet. Int.* **66**, 160–202 (1991).
3. Tanimoto, T. *J. Phys. Earth* **38**, 493–509 (1990).
4. Jarvis, G. T. & Peltier, W. R. *J. geophys. Res.* **91**, 435–451 (1986).
5. Gudmundsson, O., Davies, J. H. & Clayton, R. W. *Geophys. J. Int.* **102**, 25–43 (1990).
6. Tanimoto, T. *Geophys. J. Int.* **100**, 327–336 (1990).
7. Dziewonski, A. *J. geophys. Res.* **89**, 5929–5952 (1984).
8. Hager, B. H. et al. *Nature* **313**, 541–545 (1985).
9. Garwin, L. *Nature* **351**, 699–701 (1991).
10. Hager, B. H. & Richards, M. A. *Phil. Trans. R. Soc. A* **328**, 309–327 (1989).
11. Davies, G. F. *J. geophys. Res.* **89**, 6017–6040 (1984).
12. Gurnig, M. & Zhong, S. *Geophys. Res. Lett.* **18**, 581–584 (1991).

Actin alone in lamellipodia

Julian Heath and Bruce Holifield

LAMELLIPODIA are the thin, veil-like marginal regions of well-spread cells that demonstrate ruffling and are the primary protrusive organ of tissue cells. An innovative approach to understanding lamellipodial protrusion is described on page 126 of this issue by Theriot and Mitchison¹, who followed the behaviour of a fluorescent actin probe in a rather remarkable type of cell, the fish epidermal keratocyte.

Keratocytes move up to 30 times faster than avian and mammalian fibroblasts, the cell types most commonly used by students of the cytoskeleton and cell motility. Mammalian fibroblast lamellipodia contain a crisscross network of actin filaments interspersed by larger bundles called ribs (Fig. 1a). Lamellipodia fringe the large lamellar regions that contain the bulk of the actin cytoskeleton in the form of anchoring stress fibres and cortical arrays. By contrast, the actin cytoskeleton of fish and amphibian epidermal keratocytes, which are considerably smaller than fibroblasts, seems to be located primarily in a single anterior lamellipodium (Fig. 1b). Keratocytes can be considered as essentially lamellipodia that carry the nuclear region along behind them as baggage,

and their simplification of form and cytoskeletal organization, and high motility, make them ideal for studies on actin dynamics.

Theriot and Mitchison coupled resorufin iodoacetamide to G-actin to produce a caged compound (CR-actin) that fluoresces strongly only after activation by 360-nm light. The probe polymerized with unlabelled monomeric actin to form filaments *in vitro* and was incorporated into normal actin filament bundles *in vivo*. After microinjecting keratocytes with CR-actin, Theriot and Mitchison activated fluorescence in a bar-shaped area of the lamellipodium and imaged it by digital video microscopy over a period of time. In control cells injected with a similar but nonpolymerizing actin probe, the signal rapidly decayed as the probe diffused throughout the lamellipodium. With CR-actin, however, the fluorescent bar remained intact and stationary (with respect to the substratum), ending up at the rear as the cell moved forwards. But there was a significant decrease in intensity of the signal in less than a minute post-activation.

How do these findings relate to our knowledge of actin organization in lamellipodia?

Figure 2 shows three possible models of actin dynamics in lamellipodia. Experiments *in vitro* show that, under suitable conditions for polymerization, actin monomers add on at one end of a filament and fall off at the other (minus) end in a process called treadmilling. In fibroblastic cells, electron microscopy reveals that actin filaments span the lamellipodium with their fast-growing (plus) ends towards the tip of the lamellipodium². That actin treadmilling occurs in lamellipodia was given credence by Wang³, who introduced fluorescent actin into fibroblast lamellipodia; a spot photobleached in a non-extending lamellipodium moved rearwards with respect to the substratum, which shows actin monomers are added to filaments distally, creating a centripetal flux (Fig. 2a, b).

Curiously, a fast rearward flux of structure can also be directly seen by high-resolution video microscopy in the lamellipodia of fibroblasts and neuronal cells whether they are advancing or stationary^{4,5}. So how is actin polymerization linked to cell migration? From the absence of myosin II from

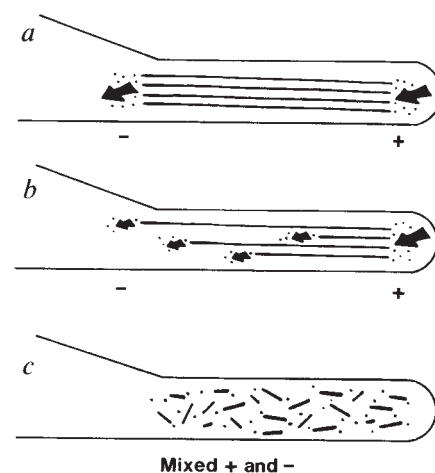


FIG. 2 Simplified schemes for actin filament dynamics and density in lamellipodia. The first model (a) embodies a polarized array of filaments of similar lengths having the fast-growing ends (+) distal and the disassembling ends (-) located at the base of the lamellipodium. This arrangement results in a net rearward flux of actin monomer with respect to the substratum. Forward extension of the lamellipodium results if the treadmilling filaments become attached to stationary substratum adhesions; flux will cease and polymerization will drive extension. In b, the filaments depolymerize throughout the width of the lamellipodium instead of only at its base, resulting in a gradient of filament density. In a forthcoming paper¹², Symons and Mitchison describe the occurrence of such a gradient in fibroblasts. Model c is the nucleation-release model, described by Theriot and Mitchison¹. Short filaments are distributed randomly with respect to polarity. Filaments assemble and disassemble throughout the lamellipodium but no net rearward flux of monomer is generated relative to the substratum.

lamellipodia it seems that protrusion does not require an actomyosin-like contraction; myosin I may be involved, however. Treadmilling could be harnessed to drive extension by the anchoring of filaments at stationary-cell-substratum-adhesion points with some kind of slippable molecular clutch⁶.

As we show here, the organization of actin filaments in keratocyte lamellipodia appears by light microscopy^{7,8} to be quite similar, if not identical, in pattern to that seen in the lamellipodium of fibroblasts (compare the images in Fig. 1). The crisscross pattern derives from fine bundles that, in fibroblasts, coalesce to form the larger ribs. Unfortunately the higher resolution of the electron microscope is still not able to resolve the question as to whether the filaments that appear to extend from tip to base in lamellipodia of fibroblasts^{9,10}, macrophages¹¹ or keratocytes⁷ are continuous or are made up of many shorter lengths.

Theriot and Mitchison show that there is a flux of fluorescent actin away from the cell margin in keratocytes, but it does not move relative to the substratum as it does in fibroblasts. This result, taken with the uniform density of actin filaments in keratocyte

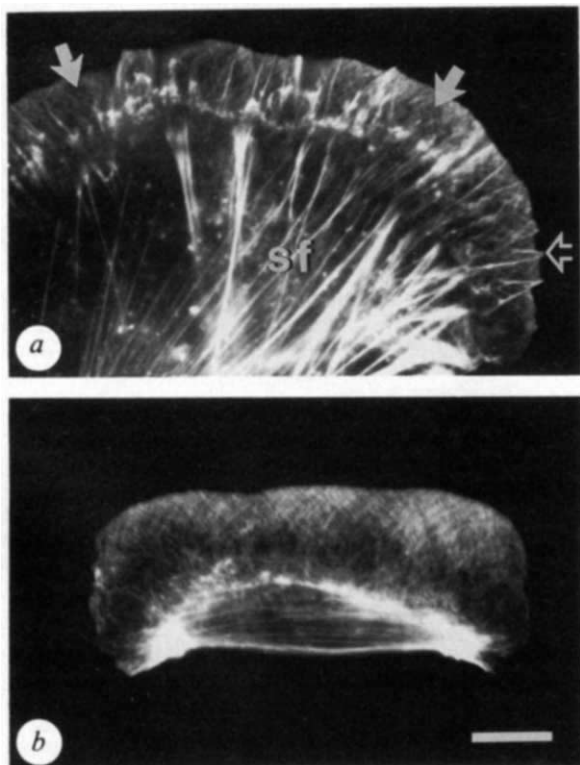


FIG. 1 Comparison of the actin cytoskeleton and lamellipodial size in a human fibroblast and a fish keratocyte. Actin filaments were stained with fluorescent phalloidin. a, The fibroblast lamellipodium contains a crisscross network of fine bundles (closed arrows) as well as larger rib-like bundles (open arrow). Note the stress fibre (sf) region. b, The keratocyte actin cytoskeleton consists of a crisscross filament network in the anterior cytoplasm. (Scale bar = 10µm.)

lamellipodia, would be compatible with model *a* (Fig. 2) in which treadmilling filaments are moved forward by a clutch mechanism. But the observed decay of the activated signal is not predicted, and so the authors theorize that there is a network of short, randomly oriented and rapidly exchanging filaments, rather than of continuous or variable length ones, in the lamellipodium. They call their hypothesis the nucleation-release model for actin filament dynamics (model *c*, Fig. 2). Presumably, the keratocyte lamellipodium advances by a greater degree of polymerization at the cell margin than at the rear.

The authors may have uncovered an important difference in actin dynamics between the rather sluggish fibroblast lamellipodium and that of the fast-moving fish keratocyte. But actin polymerization and treadmilling alone are not enough to account for cell protrusion. The secret to lamellipodial extension lies in the molecular mechanisms by which cells harness the dynamics of actin to drive the filament network forward. Keratocytes may provide us with some new details of these mechanisms.

GEOCHEMISTRY

Stability of petroleum

J. M. Hayes

PETROLEUM is much more durable in the Earth's crust than many had suspected, kinetic studies reported by Mango, on page 146 of this issue¹ and elsewhere^{2,3}, indicate. Textbooks identify thermal degradation of kerogen, the insoluble, macromolecular organic matter found in sedimentary rocks, as the nearly exclusive source of petroleum and as the main source of natural gas. But thermal cracking of petroleum has been identified as an additional source of natural gas even though evidence of this is lacking⁴. The new results indicate that this process is insignificant. In effect, estimates of the subsurface lifetimes of petroleum hydrocarbons at 150 °C have been increased from less than 50 million years^{4,5} to more than the age of the Earth, 4,500 million years.

We need not search long for the cause of this huge difference. There are two essential issues concerning organic material in sedimentary rocks: how does it get there, and by what mechanisms does it develop to yield fossil fuels? The first question relates to the synthesis, preservation and initial burial of organic material in sediments. Its dimensions are biological, ecological and oceanographic, and it can be effectively studied through observations and analyses of relatively accessible modern systems. In contrast, the second question — touched on by Mango — is chemical and geophysical. The thermal processes that alter and remobilize organic material in sedimentary rocks are occurring now, but they are difficult to observe in nature or to model in the laboratory.

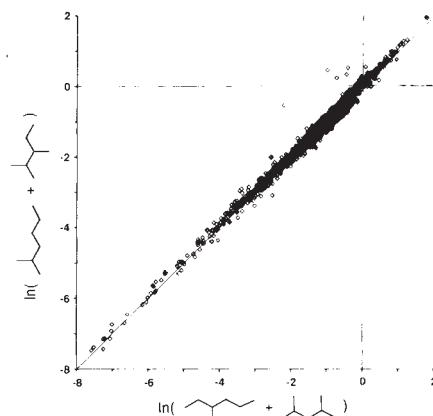
As further comparisons are made between actin filament organization and patterns of substratum adhesion in differing cell types, we should come to a better understanding of the factors controlling the locomotory behaviour of cells. □

Julian Heath and Bruce Holifield are in the Departments of Cell Biology and Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA.

1. Theriot, J. A. & Mitchison, T. J. *Nature* **352**, 126–131 (1991).
2. Small, J. V., Isenberg, G. & Celis, J. E. *Nature* **272**, 638–639 (1978).
3. Wang, Y.-L. *J. Cell Biol.* **101**, 597–602 (1985).
4. Fisher, G., Conrad, P. A., DeBasio, R. L. & Taylor, D. L. *Cell Motil. Cytoskel.* **11**, 235–247 (1988).
5. Forscher, P. & Smith, S. J. *J. Cell Biol.* **107**, 1505–1516 (1988).
6. Mitchison, T. & Kirschner, M. *Neuron* **1**, 761–772 (1988).
7. Bereiter-Hahn, J., Strohmeyer, R., Kunzenbacher, I., Beck, K. & Voth, M. *J. Cell Sci.* **52**, 289–311 (1981).
8. Cooper, M. S. & Schliwa, M. *J. Cell Biol.* **102**, 1384–1399 (1986).
9. Small, J. V. *J. Cell Biol.* **90**, 222–235 (1981).
10. Rinnerthaler, G., Herzog, M., Klappacher, M., Kunka, H. & Small, J. V. *struct. Biol.* **106**, 1–16 (1991).
11. Hartwig, J. H. & Shevlin, P. J. *J. Cell Biol.* **103**, 1007–1020 (1986).
12. Symons, M. H. & Mitchison, T. J. *J. Cell Biol.* (in the press).

They occur at depth in sedimentary formations and over time intervals that are literally geological. When laboratory conditions are chosen so that mobile hydrocarbons form at rates compatible with the completion of research projects, a 10⁸-fold compression of the time spans may be involved, and small differences in model systems or in experimental results can lead to enormous differences in geological projections.

To avoid such extrapolation, organic geochemists can seek to infer processes by



A log-log plot of the summed weight percentages of 2-methylhexane and 2,3-dimethylpentane against those of 3-methylhexane and 2,4-dimethylpentane in samples of petroleum³. Points represent more than 2,000 different oils. The compounds contributing to each axis were chosen on the basis of a postulated sequence of chemical reactions⁴.

studying the composition of petroleum itself. That is, if we can't determine how rapidly petroleum is degraded thermally, might we at least reconstruct the reaction pathway? This should not be an impossible task. It has been stressed that the chemistry of carbon is so exquisitely detailed and variable that the information content of petroleum must be great. According to this view, there are so many details in the chemical composition of petroleum that accounting for all of them will provide an abundance of theoretical constraints. But it is more to the point to observe that there are so many details that one hardly knows where to begin.

For example, all possible molecular structures with the formula C₇H₁₆ are present in petroleum. It is not by accident that considerable attention has fallen on these compounds. They are often abundant, and the range of possible structures is significant but analytically manageable. Only a few of the structures represent unrearranged fragments of biosynthetic carbon skeletons, indicating that the distribution of carbon among the nine isomers must be controlled largely by the thermal processes ultimately responsible for production of petroleum. The sequence of concentrations is quite consistent, with the straight-chain isomer being most abundant and highly branched isomers least abundant. Although it is clear that the distribution does not reflect thermodynamic equilibration³, the highly regular pattern had not been linked to a plausible, coherent chemical process.

But Mango recently showed² that the relative abundances of the four most abundant branched isomers follow the relationship summarized in the figure. The pattern of covariation is defined by points representative of all of the more than 2,000 oils in the Shell-Houston database, independent of their origin. This is, however, a double-logarithmic plot and some scatter is apparent. Significantly, linear plots reflected much tighter correlations ($r > 0.999$) when only oils from a single source bed were considered.

Mango related the scatter in the all-oils plot to varying abundances of C₇-precursor structures in kerogens and concluded³ that the systematic trends (in all oils) and the tight correlations (within subsets) would arise only if all compounds are derived from kerogen through a kinetically governed reaction network, with the fixed abundance ratios reflecting the ratios of rate constants at branch points within the network. Further study of the kinetic scheme showed that some additional abundance correlations should be found and that certain other abundance ratios should vary widely. Both predictions were fulfilled³.

This does get us closer to assessing the stability of petroleum hydrocarbons in the subsurface. A key feature of the postulated reaction network is the existence of cyclic intermediates and the generation of additional products with five- and six-membered

rings of carbon atoms. These compounds have the formula C_7H_{14} and, in nearly all petroleum samples examined, their abundances relative to those of the C_7H_{16} compounds are consistent with the prediction of Mango's kinetic scheme. On mechanistic grounds, it is expected that the alkane ring systems are more stable than the linear structures, and this is confirmed by results of Mango's differential cracking experiments^{1,6}.

If the subsurface lifetimes of hydrocarbons were as brief as suggested by earlier estimates, we would expect to find many oils in which relative abundances of cyclic structures had increased as the less stable straight-chain compounds were degraded. On the contrary, in the present report¹, Mango shows that, even though the expected enrichment of cyclic structures can be generated in

the laboratory by extreme thermal stress, it is almost never observed in nature. The structural evidence from field observations — that enrichment of cycloalkanes is nearly absent — thus comes down strongly on the side of the new laboratory evidence for long subsurface lifetimes. □

J. M. Hayes is in the Biogeochemical Laboratories, Department of Geological Sciences, Indiana University, Bloomington, Indiana 47405, USA.

1. Mango, F. D. *Nature* **352**, 146–148 (1991).
2. Mango, F. D. *Science* **273**, 514–517 (1987).
3. Mango, F. D. *Geochim. cosmochim. Acta* **54**, 1315–1323 (1990).
4. Quigley, T. M. & Mackenzie, A. S. *Nature* **333**, 549–552 (1988).
5. Barker, C. Am. Assoc. Petrol. Geol. Bull. **74**, 1254–1261 (1990).
6. Mango, F. D. *Geochim. cosmochim. Acta* **54**, 23–27 (1990).

VISUAL NEUROSCIENCE

Misaligned viewpoints

A. J. Parker

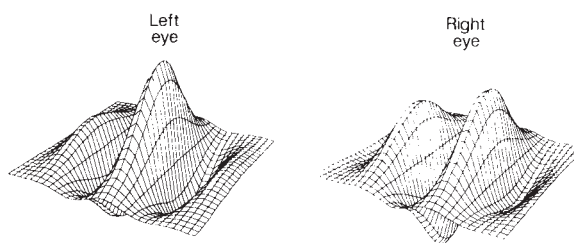
THE use of binocular vision to recover three-dimensional structure depends upon the geometrical fact that each eye acquires a slightly different view of the world. Thus, differences in depth within the scene are transformed into small differences between the images on each retina, notably in the relative horizontal positions of features¹. On page 156 of this issue², DeAngelis, Ohzawa and Freeman offer evidence that the fine structure of receptive fields in the visual cortex is shaped to exploit this information.

For the brain to exploit horizontal disparities, the correspondence between features on each separate retina must somehow be established. Physiologically, one way to do this is to have binocular detectors, which are tuned to identical properties of the image but are selectively sensitive to small geometric differences in horizontal position. Such detectors have been found³, but the present evidence indicates that the mechanisms for establishing similarity and binocular correspondence may be more diverse than has been supposed.

The new analysis provides more detail about the spatial structure of the visual receptive fields of neurons in cortical area V1 (striate cortex) of cats. This is the first stage in the visual pathway where binocular interactions are readily observed and many of the neurons can be excited to some degree by stimulation through either eye⁴. Classically, these receptive fields are known to be restricted in spatial extent, selective for the orientation of image features⁴ and sensitive

to only a limited range of spatial frequencies within the image⁵.

These findings are captured by a model in which the receptive field is described by a two-dimensional sinusoidal wave packet



Two idealized receptive fields of cortical simple cells that are described by a two-dimensional gaussian multiplied by a (co)sinusoidal wave (Gabor function⁷). These receptive fields are selective for orientation and spatial frequency. The two fields represent the spatial profiles of a binocular cortical cell as measured through the left and right eyes. The phase difference (90°) between sinusoids describing the left and right eyes' receptive fields allows the binocular signal from the cell to be sensitive to stereoscopic disparity.

(sinusoid multiplied by a two-dimensional gaussian, known as a Gabor function — see figure)^{6,7}. The detailed shape of the receptive field is determined by several parameters in the model, but the new evidence suggests that the phase of the sinusoid may be crucial for the disparity selectivity of a particular population of cortical cells. For some binocular cortical cells, the inference is that, when receptive fields are measured through the left and right eyes, the measurements are identical except in this parameter.

A shift in the phase, leaving other parameters constant, changes the locations of the peaks and troughs of the receptive fields in the left and right eyes (see figure). This would give the cell some sensitivity to binocular disparity and, indeed, it is possible to

express binocular disparities as shifts in the spatial phase of a sinusoidal signal, in much the same way that interaural time differences in the auditory system can be expressed as a temporal phase lag.

The scheme proposed by DeAngelis *et al.*² bears upon the general problem of phase recovery in signal processing. Indeed, there are specific computational proposals for treating stereoscopic disparities as if they were spatial phase differences⁸. Similarly, there are proposals for the use of quadrature pairs of filters in the $x-t$ spatiotemporal domain for the detection of visual motion^{9,10}.

Now that the new conceptual scheme has been put forward, it is important to understand its relationship with earlier results³, including those of Freeman's group¹¹. That work used truly binocular visual stimulation and it would be good to see the reverse correlation technique^{2,11} applied binocularly. The present measures have provided no information on the possibility of interocular differences in the location of the peak of the gaussian envelope (see Fig. 1a of the paper on page 157). Yet the phase differences observed in this work cannot be translated into stereoscopic disparity without knowing the location of the envelope peak as well.

It would be ironic if the position of the envelope peak were found to covary with the phase difference, but such an organization could actually be produced by a developmental mechanism that attempts to construct binocular receptive fields whose left and right eye inputs are as closely matched as possible, thus presumably causing the cell to be tuned to zero disparity. If the initial distribution of the influences of the left and right eyes were misaligned, then phase differences could be induced by a developmental mechanism that refines binocular connections based on the degree of binocularly correlated (visually driven) activity. Such an outcome would return the cell towards being tuned to zero disparity, and indeed it is a puzzle to understand what kind of developmental process could lead to the appearance of controlled degrees of *dissimilarity* between receptive field structures for the left and right eyes. □

A. J. Parker is in the University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK.

1. Wheatstone, C. *Phil. Trans. R. Soc.* **128**, 371–394 (1838).
2. DeAngelis, G.C., Ohzawa, I. & Freeman, R.D. *Nature* **352**, 156–159 (1991).
3. Barlow, H.B., Blakemore, C.B. & Pettigrew, J.D. *J. Physiol.* **193**, 327–342 (1967).
4. Hubel, D.H. & Wiesel, T.N. *J. Physiol.* **160**, 106–154 (1962).
5. Cooper, G.F. & Robson, J. *IEE Conf. Pub.* **42** (IEE/NPL Conf. on Pattern Recognition, 1968).
6. Daugman, J. *Vision Res.* **20**, 847–856 (1980).
7. Marčelja, S. *J. Opt. Soc. Am.* **70**, 1297–1300 (1980).
8. Jenkin, M.R. *Tech. Rep. RBCV-TR-88-22* (Dept Computer Sci., Univ. Toronto, 1988).
9. Watson, A.B. & Ahumada, A. *J. Opt. Soc. Am.* **2**, 322–342 (1985).
10. Adelson, E.H. & Bergen, J.R. *J. Opt. Soc. Am.* **2**, 284–299 (1985).
11. Freeman, R.D. & Ohzawa, I. *Vision Res.* **30**, 1661–1676 (1990).

Supersalt

It is alleged that a poached egg, sprinkled with ATP, will contract; this is not to say that a poached egg resembles a muscle but rather that a highly charged ion can exercise large electrostatic effects of a non-specific kind. A striking demonstration of this principle now comes from Y. Goto *et al.* (*J. Biochem. (Tokyo)* **109**, 746–750; 1991). Acidified (therefore cationic) apomyoglobin and cytochrome *c* are unfolded, but enter a compact, structured molten-globule state when salt is added. With ATP, the transition has a mid-point below 1 mM. With adenosine tetraphosphate, it is an order of magnitude down, while ADP pushes it up by a similar factor. A Lys–Leu copolymer will undergo a conformational transition at neutral pH at sub-millimolar ATP concentrations. A powerful charge-condensation phenomenon is evidently at work. Those who work histones, protamines and the like should take note.

Young ones

LARGER than Jupiter, smaller than the Sun, brown dwarfs are failed stars that may be numerous enough to account for much of the Galaxy's mass. But, despite some false claims, none has been observed. The answer, according to G. S. Stringfellow (*Astrophys. J.* **375**, L21–L25; 1991), is "to catch them while they are hot and young". Although too light to burn through nuclear fusion as do normal stars, brown dwarfs retain the heat from their initial collapse and radiate faintly in the infrared. For old dwarfs, this glow is pretty indistinguishable from that of feeble, very-low-mass stars. But the distinction is sufficient when the objects are newly formed (less than 100 million years old). The place to look, says Stringfellow, is in young star clusters, such as the Pleiades, where all the objects are of a similar age and the most can be made of the distinction.

Toxin two

THOSE in pursuit of a safe vaccine against cholera may now have to take into account a newly discovered second toxin produced by the causative bacterium, *Vibrio cholerae* (A. Fasano *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 5242–5246; 1991). Vaccines based on live attenuated strains should be effective if they are deleted in the gene encoding cholera toxin, the substance believed to be responsible for the intestinal damage that characterizes cholera. But Fasano *et al.* now show that strains negative for cholera toxin produce a protein which disrupts the adherent cellular junctions between the cells of the small intestine, and which can still produce mild diarrhoea in volunteers.

SQUIDS in with applications

John Clarke

SINCE specialists working on superconducting devices last met in Berlin, in 1985, the discovery of the new family of copper-oxide-based, high-temperature superconductors has given great impetus to the field. Indeed, as became clear at the subsequent meeting last month*, progress towards making devices based on these materials and operating in liquid nitrogen at 77 K has been faster than many had expected. Nevertheless, it is devices based on conventional, low-temperature materials that continue to be used in practical applications and that are available in increasingly sophisticated, multichannel systems.

A unifying element for many of the participants was the Josephson junction, a weak link between two superconductors through which supercurrents tunnel quantum mechanically. Hope seems to be fading (Y. Okabe, Tokyo) for high-temperature Josephson junctions with the hysteretic current-voltage characteristics displayed by their low-temperature counterparts and used, for example, in computer logic and memory elements and in the voltage-frequency standard. But there has been much progress towards making nonhysteretic junctions from $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (YBCO) — just the kind needed for superconducting quantum interference devices (SQUIDs). When the d.c. SQUID — two Josephson junctions connected in parallel on a superconducting loop — is appropriately current-biased, the voltage across it depends periodically on the magnetic flux applied to the loop. The period is the flux quantum, $\phi_0 = h/2e$ (h is Planck's constant and e , the electron charge); with the aid of an electronic feedback circuit one can detect changes in flux as small as $10^{-5} \phi_0$.

In the first thin-film high-temperature SQUIDs, the junctions were the naturally occurring boundaries between grains with different crystalline orientations. Today, although many junction types have been investigated, the most successful devices operating at 77 K are based on the controlled growth of grain boundaries. In one method R. H. Koch, IBM Yorktown Heights, M. Kawasaki, IBM) a SrTiO_3 substrate is cut and fused to produce a bicrystal with a known misorientation across the interface. A film of YBCO grown with its *c*-axis perpendicular to the bicrystal replicates the grain boundary and, patterned into a narrow strip, produces a junction at a prescribed location. F. C. Wellstood (University of California, Berkeley in collaboration with Conductus Inc., Sunnyvale) described a 'bi-epitaxial' process in which a SrTiO_3 buffer layer followed by a YBCO film are deposited on a sapphire substrate on which an MgO seed

layer has previously been grown and patterned. This process yields a 45° grain boundary at the edges of the MgO, enabling one to place junctions at will.

SQUIDs are very sensitive to magnetic flux, but, especially with the small effective area allowable for thin-film high-temperature devices, not particularly sensitive to magnetic field (flux divided by area). To make a useful magnetometer one has to couple the SQUID to a thin-film 'flux transformer' — a multiturn coil with dimensions to match the SQUID (say 0.25 mm across) connected to a much larger (say 10 mm) single-turn pick-up loop. Wellstood described our 'flip-chip' magnetometer in which the SQUID and flux transformer are on separate chips that are pressed together. At 77 K, the sensitivity is sufficient to obtain a magnetocardiogram.

Koch also reported flip-chip devices, including a planar gradiometer in which there are two pick-up loops of opposite sense making the flux transformer nearly insensitive to uniform fields. However, flip-chips may soon be *passé*. We (the Conductus/Berkeley group) reported an integrated magnetometer in which the flux transformer and SQUID are deposited on a single chip. The device involves seven epitaxial layers, three superconducting and four insulating, indicating that relatively sophisticated high-temperature circuits may not be far away. Nevertheless, the low-frequency excess noise remains distressingly high in high-temperature SQUIDs. This noise is a significant limitation for biomagnetic applications, for which high sensitivity at frequencies around 1 hertz is required.

The moment of truth for biomagnetic applications of low-temperature SQUIDs may already be at hand. Long regarded as the application most likely to have a substantial commercial impact, magnetic measurements of heart and brain function have been slow to gain clinical acceptance. The main drawback, until recently, has been the absence of the multichannel systems necessary for images of magnetic sources with useful spatial resolution. But no fewer than eight groups — four of them commercial — now produce magnetometers with 24–38 channels, all based on niobium, which has become the preferred low-temperature material. S. Schneider (Siemens) described a study of 20 epileptic subjects in whom the epileptic centre was located magnetically. Of these, 17 were treated with a γ -ray knife and 14 were cured.

Needless to say, SQUIDs were by no means the only practical devices of interest. A high-temperature flux-flow transistor has been devised with a gain and sensitivity that, at microwave frequencies, are competitive with semiconducting devices (Martens, Los

* SQUID '91, Berlin, 18–21 June 1991.

Alamos National Laboratory). Perhaps the most elegant physics, however, was in submicrometre junctions and in junction arrays. These experiments are concerned with the passage of single electrons through tunnel junctions with capacitances of a femtofarad or less operating at millikelvin temperatures (see T. A. Fulton *Nature* **346**, 408; 1990). By synchronizing the rate of flow of these electrons with a radio-frequency signal f , one endeavours to produce steps of constant current $I = nef$ (n is an integer). Considerable ingenuity is needed to defeat parasitic capacitance: solutions tend to involve multiple junctions, for example, the electron pump (C. Urbina, Saclay), the turnstile (H. Mooij, Delft University) and the array turnstile (P.

Delsing, Chalmers University, Göteborg). So far the current–frequency relation is accurate only to about 1 part in 10^3 , whereas 1 part in 10^7 or more is needed for a standard. The motivation to improve is, however, very high: a current–frequency standard would enable one to link the Josephson voltage–frequency standard and the quantum Hall effect current–voltage standard, thus testing these very fundamental relationships with great precision. Perhaps at the next SQUID meeting (in Berlin, 4 years hence?) we shall hear the outcome. □

John Clarke is at the Department of Physics and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA.

PALAEOANTHROPOLOGY

Teeth, sex and species

Lawrence Martin

ONE person's evidence for extreme sexual dimorphism is another's evidence for species diversity. Thus, on page 151 of this issue¹, Kelley and Xu argue that fossil apes from the late Miocene site at Lufeng, China, dating to about 8 million years ago, demonstrate an extreme level of sexual dimorphism in dental dimensions that exceeds the levels in the most sexually dimorphic living apes, the orang utan and gorilla (see figure). This indeed is the inescapable conclusion to be drawn from viewing the Lufeng assemblage as a single species. But in an alternative interpretation of the metric variation of the same material, Oxnard² has contended that two species are present, and that the extremely low level of sexual dimorphism in at least the smaller species, *Ramapithecus lufengensis*, indicates that this form is more closely related to humans than is the larger form. Such different views of a single fossil sample illustrate a basic problem in the interpretation of fossil material, the delineation of species in the fossil record.

Species are, for most biologists, the fundamental natural group of organisms. Systematists may argue about the level of variation that denotes a tribe rather than a family (for example Hominidae against Hominini as the exclusive group for humans and their bipedal fossil relatives), but, by and large, some variant of a biological species concept works pretty well for most life forms. It is natural, therefore, to wish to apply biologically meaningful limits to the delineation of species in the fossil record. This is not simply a question of having stamps pasted in on the right page of the album, but is central to all attempts to reconstruct phylogeny and to interpret the behaviour of extinct organisms.

Following the work of Gingerich³, most such analyses have been based on comparisons of coefficients of variation in fossil samples to those in extant species. Some workers, such as Martin and Andrews^{4,5},

have argued for the use of range-based statistics in addition. Recent work by Cope⁶ has shown that although, as Simpson⁷ pointed out, such measures often fail to recognize the presence of several taxa in a fossil sample because range estimates are so dependent on sample size, they provide strong, if not defin-



Skulls of a male (left) and female gorilla. The level of sexual dimorphism in this species, here principally seen in relative skull and canine size, is among the highest seen in extant primates.

itive, evidence for the presence of more than one species when their values exceed comparative levels. Fossil samples are very different from modern samples in being primarily limited to teeth, jaws and bones, or fragments of them, for which gender can only be inferred. Fossils do not have fur or feathers, do not make calls and, most importantly, do not breed. In life, such samples may not have occurred together in space or time (that is, may not have been sympatric or synchronous), and fossil assemblages may be built up over long periods of time, which may increase levels of variation.

The Lufeng apes were first described as belonging to two species, *Ramapithecus*

*lufengensis*⁸ and *Sivapithecus yunnanensis*⁹. Later analyses have sought to determine whether the level of variation in the sample is compatible with the interpretation that a single species is represented, based on levels of variation seen in modern species of hominoids^{2,5,10–16}. Wu and Oxnard^{10,11} and Oxnard² found levels and patterns of variation in excess of their expectations for a single species.

But Wu has since come to favour the single-species view, the variation being explained as the result of a high degree of sexual dimorphism^{12,13}. Kelley and Etler¹⁴ and Wood and Xu¹⁵ have employed coefficients of variation, and found several tooth measurements in which the level of variation at Lufeng exceeds the maximum value encountered in modern apes. In spite of this, both sets of authors concluded that the Lufeng sample was compatible with a single-species interpretation, citing a variety of factors to explain the excess variation. In an analysis of range-based statistics¹⁷, Martin and Andrews⁵ found that the Lufeng sample exceeds the maximum value found in modern hominoids for 21 out of 32 dental measurements. Without special pleading, the single-species interpretation of Lufeng may be considered to have been falsified by these comparative studies of variation.

Kelley and Xu¹ have now sexed individual specimens in order to quantify levels of sexual dimorphism that were previously only implicit. They agree with others that a single species is present^{12–15}, but do not agree that this sample shows a pattern comparable with that of modern species. Rather they find that molar sexual dimorphism in the Lufeng sample exceeds that in any modern ape and conclude that patterns and levels of within-species variation in the past differed from those seen today.

Cope⁶ working on sympatric guenons (*Cercopithecus monkeys*), and Plavcan¹⁸, working on a variety of catarrhine primates, have shown that multiple-taxon samples of modern

species can exhibit levels of variation comparable with, and thus difficult to distinguish from, single species. If such samples were encountered in the fossil record they would appear to be a single species. So a single species can never be demonstrated to exist in a fossil sample; rather, it is merely the null hypothesis to be falsified. Most researchers feel that the null hypothesis can be tested by studies of the levels of variation in fossil samples against those seen in extant species. Kelley¹⁶, however, argues that this approach represents misapplication of the concept of uniformitarianism, one that constrains past life forms to have existed within modern limits (see ref. 19). His view in turn raises

very real problems as to how to decide upon what, in a particular instance, constitutes a species.

A particular advance described by Kelley and Xu¹ is determination of the sex of individual, isolated canine teeth using indices of crown height and tooth length. They show that all of the larger canine teeth in the Lufeng sample are from males and that all of the small ones are from females. If this analysis holds up it demolishes Wu and Oxnard's^{2,10,11} interpretation of two species distinguished by size, as that model would predict the presence of both sexes in each of the size groups.

In dealing with morphologically homogeneous, but metrically variable samples, two positions have traditionally been adopted — that the sample represents a single, highly dimorphic species; or that it represents two species, each of which displays a relatively low level of dimorphism. But there is a third possibility, which is that the sample represents two dimorphic species whose size ranges overlap. Cope⁶ has shown that such could easily be the case with guenons — teeth of two or three species of sympatric *Cercopithecus* that are morphologically homogeneous and overlap in size can be mixed to provide a sample in which two clearly defined size clusters of canine teeth occur, one male and one female, each of which contains a single sex sample of more than one species. So the Lufeng pattern of variation can be mimicked in a multiple-species sample of *Cercopithecus*.

The question that palaeoprimatologists must confront is whether the guenon example is relevant to fossil apes. Kelley¹⁶ feels that it is not, and that only hominoids provide a suitable model for the patterns of geographical distribution and variation. Hominoids however exhibit low species diversity compared to other groups of primates and have few sympatric taxa that can be used to model what might happen in a fossil assemblage. Kelley believes that hominoids of similar morphology and similar size do not occur sympatrically because aspects of ape biology would prohibit it. But we know that this kind of sympatry occurs in guenons, and also in lemurs²⁰. Thus, the argument that present hominoid distribution and variation reflects ecological limits requires further analysis before it is accepted as the basis on which we interpret the past. This is particularly the case as most workers recognize the hominoid primates as relict populations. Hominoid primates were once much more speciose and widely distributed than is the case today, and it is well known that species diversity in relict faunas is impoverished. Perhaps modern hominoids are too impoverished to provide a suitable yardstick against which one can compare fossil samples from a time when hominoids were more diverse and more widely distributed.

One might look to the fossil record to resolve this conundrum. But samples of Mio-

cene apes that many workers interpret as containing more than one species, and thus in support of the view that modern apes are impoverished in species diversity (for example *Dryopithecus* fossils in the Miocene of Spain, fossil apes from the middle Miocene of Paşalar, Turkey¹⁷, *Sivapithecus* from Indo-Pakistan, and *Proconsul*²¹ from the early Miocene of Kenya) can also be interpreted as single species if Kelley's position is adopted. One's analytical models (and, ultimately, one's inferences) are thus determined by one's perception of hominoid ecology past and present.

If one adheres, as I do, to the principle that hypotheses in science must be falsifiable, then, based on the criteria developed by Cope⁶, the Lufeng material is not a single species. Kelley, for one, believes however that falsifiability is not in most cases applicable in the strict sense to historical science such as palaeontology. Likewise, if one believes that morphologically and metrically similar apes are unlikely to occur sympatrically, then one will necessarily interpret all such accumulations in the fossil record as a single taxon. But the circularity of the reasoning prevents the result from corroborating the assumption.

There is no question that Kelley and Xu's work, particularly by adding gender to the equation, has advanced the analysis of palaeontological variation, at least for fossil apes. They could well be right that the Lufeng material represents a single species whose sexual dimorphism extends the limits seen in modern animals. The difficulty is that I can see no method by which we can ever find out whether they are correct or not. □

Lawrence Martin is in the Departments of Anthropology and Anatomical Sciences, State University of New York, Stony Brook, New York 11794, USA.

- Kelley, J. & Xu, Q. *Nature* **352**, 151-153 (1991).
- Oxnard, C. *Fossils, Teeth and Sex New Perspectives on Human Evolution* (Washington Univ. Press, Seattle, 1987).
- Gingerich, P.D. *J. Paleont.* **48**, 895-903 (1974).
- Martin, L. & Andrews, P. *Cour. Forsch. Inst. Senckenberg* **69**, 25-40 (1984).
- Martin, L. & Andrews, P. in *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W.H. & Martin, L.B.) (Plenum, New York, in the press).
- Cope, D. thesis, Univ. Texas, Austin (1989).
- Simpson, G. G. *Am. J. Sci.* **239**, 785-804 (1941).
- Xu, Q., Lu, Q., Pan, Y., Qi, G.-Q., Zhang, X. & Zheng, L. *Kexue Tongbao* **23**, 554-556 (1978).
- Xu, Q. & Lu, Q. *Vert. Palasiatica* **17**, 1-15 (1979).
- Wu, R. & Oxnard, C. E. *Nature* **306**, 258-260 (1983).
- Wu, R. & Oxnard, C. E. *Am. J. Primatol.* **5**, 303-344 (1983).
- Wu, R., Xu, Q. & Lu, Q. *Acta anthrop. Sinica* **5**, 1-30 (1986).
- Wu, R. *Acta anthrop. Sinica* **6**, 265-271 (1987).
- Kelley, J. & Etlar, D. *Am. J. Primatol.* **18**, 15-34 (1989).
- Wood, B. A. & Xu, Q. *J. hum. Evol.* **20**, 291-311 (1991).
- Kelley, J. in *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W.H. & Martin, L.B.) (Plenum, New York, in the press).
- Alpagut, B., Andrews, P. & Martin, L. *J. hum. Evol.* **19**, 397-422 (1990).
- Plavcan, J.M. thesis, Duke Univ. (1991).
- Gould, S.J. *Am. J. Sci.* **263**, 223-228 (1965).
- Tattersall, I. in *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W.H. & Martin, L.B.) (Plenum, New York, in the press).
- Ruff, C. B., Walker, A. & Teaford, M. F. *J. hum. Evol.* **18**, 515-536 (1989).

Prestidigitation

COMPUTERS still can't read. Even with the error-prone assistance of scanners and character-recognition programmes, the best way of getting text — and certainly handwriting — into a computer is to type it in manually. Daedalus is now changing all that. He presents DREADCO's Digital Paper, and its matching Digital Ink.

Digital Paper is hydrophobic, except for an invisible pattern of wettable dots spaced to the dot-matrix printer standards. Digital Ink can wet only the dots. Any text or image it forms on the paper is thus perfectly dot-digitized at standard spacing. More cunning still, each sheet of Digital Paper carries a matching grid of vertical and horizontal optical fibres just below the surface. Each dot is on the intersection of two fibres.

Digital Ink is a fast-setting resin with the same refractive index as the optical fibres, loaded with dark pigment-particles. When it lands on a dot, it glues the intersecting fibres together by a path short enough to contain few or no absorbing particles. Light shone down the vertical fibre is then efficiently coupled into the horizontal one.

A sheet of Digital Paper can thus be read in two ways. The human eye just sees a normal image, sharpened and clarified by dot digitization. But the special DREADCO reading-frame, with an array of light-emitting diodes along the top of the paper and an array of sensing ones down the side, scans it digitally. Each emitter diode in turn sends a brief pulse down its vertical fibre, while the sensing diodes record which horizontal fibres light up in response. Every dot on the whole sheet is digitally located by its vertical and horizontal coordinates.

Thus paper becomes machine-readable at last. Any image on Digital Paper, text and graphics together, can be instantly loaded into a computer. Its flawless standard digitization permits rapid conversion of text to characters. The whole image can be manipulated, stored or transmitted in digital form, reprinted at local or remote sites, all with no degradation.

Even full-colour print-images could be digitized. Up to seven mutually immiscible liquids are known to chemistry. So three different kinds of dot, each wettable only by ink of the right colour, should be quite feasible. A colour picture printed this way would be digitized like a television picture: indeed, Daedalus plans to match his paper exactly to television standards. One standard format of full-colour graphics will then link all the media in a wonderful rainbow alliance. Captured by video, manipulated by computer, lifted from or printed onto paper, the colour-supplement dream of endless trendy imagery will riot around us with even greater exuberance.

David Jones

Oceans of garbage

SIR — Ducie Atoll in the south Pacific (24°40' S, 124°47' W) is one of the most remote islands in the world. It is 293 miles away from the nearest inhabited island (Pitcairn Island, population around 50), and more than 3,000 miles away from the nearest continent.

As part of the Sir Peter Scott commemorative expedition to the Pitcairn Islands I visited Ducie at the end of the March and surveyed the seaward shore of Acadia, the largest island in the atoll, for pieces of non-biodegradable flotsam of human origin. The survey was a belt transect, 1.5 miles long: I walked along the shore and recorded the objects I saw. The list is not complete: obviously the method is biased against smaller objects which were less visible from a distance. In total I recorded 953 objects, listed in the table.

Ducie is rarely visited: it has little to offer to passing boats (of which there are probably about 30 yachts per year *en route* between Easter Island and French Polynesia). There is no standing fresh water on the island, and the one species of plant found there is not edible. Pieces of pumice and tree-fern trunks attest to the ability of objects to float from South America, but most flotsam is likely to come from ships throwing their rubbish overboard.

On nearby Henderson Island (190 miles west of Ducie) I found equally bizarre flotsam on the beaches. In more than 3 months beachcombing, I found enough toy soldiers

(six) and vehicles (two jeeps and a tank) to make a toy army, as well as such diverse items as full bottles of body lotion and a medical syringe with an attached needle.

If so much rubbish is washed ashore on these small and extremely isolated islands, it makes one wonder just how much more is still floating on the surface of the oceans.

TIM BENTON

Department of Zoology,
University of Cambridge,
Downing Street,
Cambridge CB2 3EJ, UK

Ca²⁺ channels or voltage sensors?

SIR — There is considerable confusion about the precise role and possible diversity of dihydropyridine receptors (DHPRs) in skeletal muscle: at least some of these receptors must act as voltage-dependent calcium channels in the surface membrane^{1,2} whereas others act as voltage-sensors controlling the calcium release channel/foot structures in sarcoplasmic reticulum^{3,4}. It has been suggested that only a small fraction of DHPRs are 'functional' calcium channels¹, but this may be true only at a given instant and all these receptors may perhaps act as calcium channels at some time². Thus the possibilities are that: (1) there are two fundamentally different types of DHPR, a few being calcium channels and almost all being voltage-sensors which are inherently unable to act as calcium channels⁵; (2) all DHPRs can be calcium channels, with some local factor *in vivo* dictating that each is either a voltage-sensor or a calcium channel, either permanently or at a particular instant; or (3) a DHPR can simultaneously have both functions².

In support of (1), Catterall and colleagues have suggested³ that only few (2–3%) of the DHPRs have the complete *M*_v 212,000 (212K) sequence and can act as calcium channels, whereas most have had part of their C-terminal end removed, reducing them to 175K proteins which can act only as voltage-sensors and not as calcium channels. (This would mean that all previous biochemical studies were incorrect in attributing calcium channel activity to the 175K form, and also does not explain a similar difference between the measured and predicted size of the receptor in cardiac muscle.) However, Catterall's group also found⁶ that phosphorylation of the DHPRs increased the number of 'functional' calcium channels incorporated into lipid vesicles at least nine times, implying that many if not all of the receptors function as calcium channels (at least 36–54%, assuming that only half of the receptors were 'inside-out' and could be phosphorylated), thus eliminating hypothesis (1).

A further insight comes from experiments involving the expression of cardiac and skeletal muscle DHPRs in dysgenic skeletal muscle^{7,8}, in which a similar ratio of calcium channel conductance to charge movement is found for both types of receptor (36 and 85 nS pC⁻¹ for expressed and native skeletal DHPRs and 55 nS pC⁻¹ for expressed cardiac DHPRs). Assuming that the non-linear charge movement indicates the number of receptors present, Adams *et al.* suggest that "the two types of DHPR do not appreciably differ in their ability to function as channels"⁸. Noting that all cardiac DHPRs are usually thought to be calcium channels, this seems to support hypothesis (3). However, consideration of the absolute ratios shows that the situation is more complicated than it first appears: the above ratios are virtually the same as that found previously in adult skeletal muscle² (about 50 nS pC⁻¹ in three different muscle types, using the nifedipine-sensitive charge movement and allowing for 10 mM Ca²⁺), but are four to eight times lower than that for calcium channels in cardiac cells⁹ (> 300 nS pC⁻¹ for 10 mM free Ca²⁺). Noting also that the number of skeletal muscle DHPRs expressed in non-skeletal muscle cells agrees well with the apparent number of functional calcium channels¹⁰, it appears that when either cardiac or skeletal muscle DHPRs are expressed in a skeletal muscle cell, the current per receptor is much lower than either when cardiac receptors are expressed in a cardiac cell or when skeletal muscle receptors are expressed in a non-skeletal muscle cell.

The intriguing implication of this result is that some factor (or absence of some factor) associated with a skeletal muscle cell appears to reduce the average current through both the cardiac and the skeletal DHPR, either by lowering the conductance or the average open-time of the channels, the latter including the case where certain DHPRs often open as calcium channels and others never do so. Perhaps this factor involves the level of phosphorylation or the association of the DHPRs with the release channels. Nevertheless, this important observation still does not allow us to discriminate between hypotheses (2) and (3).

G. D. LAMB

Department of Zoology,
La Trobe University,
Bundoora, Melbourne 3083,
Australia

- Schwartz, L. M., McCleskey, E. W. & Almers, W. *Nature* **314**, 747–751 (1985).
- Lamb, G. D. & Walsh, T. J. *J. Physiol., Lond.* **393**, 595–617 (1987).
- Rios, E. & Brum, G. *Nature* **325**, 717–720 (1987).
- Tanabe, T. *et al.* *Nature* **336**, 134–139 (1988).
- De Jongh, K. S., Merrick, D. K. & Catterall, W. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8585–8589 (1989).
- Nunkio, K., Florio, V. & Catterall, W. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6816–6820 (1989).
- Tanabe, T. *et al.* *Nature* **346**, 567–569 (1990).
- Adams, B. A. *et al.* *Nature* **346**, 569–572 (1990).
- Field, A. C., Hill, C. & Lamb, G. D. *J. Physiol., Lond.* **406**, 277–297 (1988).
- Perez-Reyes, E. *et al.* *Nature* **340**, 233–236 (1989).

HARVEST FROM DUCIE ATOLL

Buoys: large (usually 30 cm diameter)	46
small (usually 20 cm diameter)	67
pieces (glass, plastic, polystyrene)	66
Crates (bread, bottle)	14
Plastic bottles (drinks, toiletries)	71
Glass bottles (produce of > 15 countries, 35% once contained whisky)	171
Jars	18
Unidentified (or broken) plastic pieces	268
Bottle tops	74
Segments of plastic pipe	29
Pieces of rope	44
Shoes	25
Fluorescent tubes	6
Light bulbs	6
Aerosol cans	7
Food/drink cans	7
Biro tops	2
Jerry cans (all holed)	4
Gloves (1 pair)	2
Tinned meat pie (leaking, but intact)	1
Cigarette lighters (non functional)	3
Doll's heads (1 male, 1 female)	2
Copper sheeting from hulls of wrecks	8
Lorry tyre	1
Plastic skittle	1
Glue syringe	1
Small gas cylinder	1
Construction-worker's hat (brown)	1
Plastic coat hanger	1
Toy soldier	1
Toy aeroplane	0.5
Tea strainer	1
Football (punctured)	1
Plastic foot mat from car	1
Asthma inhaler	1

Chirality and cold origin of life

SIR — Recent comments on the possible non-racemic nature of extraterrestrial amino acids^{1,2} have implications for the origin of life. Two scenarios have been postulated: a terrestrial origin (the Oparin-Haldane model) and an extraterrestrial origin (initiated by Svante Arrhenius's 'panspermia' hypothesis). Neither of these, however, provides an explanation of the chiral purity of the terrestrial biosphere.

We have previously proposed an extraterrestrial origin which we term the 'cold prehistory of life'. This is based on the phenomenon of molecular tunnelling^{3,4} — the existence of a finite limiting chemical reaction rate at low temperatures resulting from quantum tunnelling^{5,6}. Experiments on formaldehyde polymerization³ provide evidence for such a process. Recently we reached three conclusions concerning the role of chirality in any model⁷:

(1) The chiral purity observed in the biosphere was achieved at the stage of prebiotic evolution and was a necessary condition for the subsequent development of self-replication.

(2) Chiral purity resulted not from the gradual (evolutionary) accumulation of an enantiomeric excess, but by a process in which mirror symmetry was broken spontaneously^{8,9}.

(3) The sign of chiral purity in the bio-organic world (L-amino acids and D-sugars) is random^{10,11} rather than predetermined by some 'global advantage factor' (for example, by the non-conservation of parity in weak interactions, particularly weak neutral currents¹²).

Spontaneous racemization at low temperatures by D $\rightleftharpoons$ L tunnelling is predicted by Hund's paradox (that is, the fact that the symmetric ground state for a chiral species is the racemic mixture). But in the 'cold prehistory of life' it is possible to inhibit D $\rightleftharpoons$ L conversion^{13,14} even at temperatures sufficient for cryochemical reactions to occur, such as those found in space.

The study by Engel *et al.*¹ on the stereoisomer and isotope (¹³C) composition of amino acids in the Murchison meteorite is of great interest in this regard. Engel *et al.* reported enantiomeric excesses of the L-isomer for alanine (D/L ratio of 0.85 $\pm$ 0.03) and glutamic acid (0.54), accompanied by a ¹³C enrichment typical of extraterrestrial organic materials (up to 30 ‰). They concluded that optically active (or rather, deracemized) compounds were present in the early Solar System.

This conclusion is consistent with point (1) above. Moreover, if this partially deracemized state represents an intermediate stage in the transition to a chirally pure state, the results can provide an indication of the timescale required for the symmetry-breaking process in point (2): according to

ref. 15 it could be as short as 10⁶–10⁷ years, whereas the whole prebiotic stage of the Earth's history lasted 2 $\times$ 10⁸–5 $\times$ 10⁸ years.

In view of the totality of data that led us to our conclusions (2) and (3), the possibility mentioned in ref. 1 that β -decay of ¹⁴C could provide a mechanism for deracemization of amino acids in meteorites must, in our opinion, be excluded.

Finally, in regard to the question of whether extraterrestrial amino acids can be delivered intact to the Earth by a large impactor^{2,16}, we note that shock waves with amplitudes of as much as 500 kbar need not destroy amino acids but can instead initiate their condensation into oligopeptides¹⁷.

VITALII I. GOLDANSKII

VLADIMIR V. KUZMIN

N. N. Semenov Institute of Chemical Physics,
USSR Academy of Sciences,
Ulitsa Kosygina 4,
117 334 Moscow, USSR

- Engel, M. N., Macko S. A. & Silfer J. A. *Nature* **348**, 47 (1990).
- Chyba, C. F. *Nature* **348**, 113 (1990).
- Goldanskii, V. I., Frank-Kamenetskii, M. D. & Barkalov, I. M. *Science* **182**, 1344 (1973).
- Goldanskii, V. I. *Nature* **268**, 612 (1977); **269**, 583 (1977); **279**, 109 (1979).
- Bell, R. P. *Proc. R. Soc. A* **139**, 466 (1933); **148**, 241 (1935); **158**, 128 (1937).
- Goldanskii, V. I. *Doklady Akad. Nauk SSSR* **124**, 1261 (1959); **127**, 1037 (1959).
- Goldanskii, V. I. & Kuzmin, V. V. *Sov. Phys. Uspekhi* **32**, 1 (1989).
- Frank, F. *Biochim. biophys. Acta* **11**, 459 (1953).
- Moroso, L. L. *Origins of Life* **9**, 187 (1979).
- Avetisov, V. A., Kuzmin, V. V. & Anikin, S. A. *Chem. Phys.* **112**, 179 (1987).
- Grossmann, S. & Mikhailov, A. S. *Z. Phys.* **B78**, 1 (1990).
- Kondepudi, D. K. & Nelson, G. W. *Nature* **143**, 438 (1985).
- Harris, R. A. & Stodolsky, L. *Phys. Lett.* **B78**, 313 (1978); **116**, 464 (1982).
- Berlin, Yu. A. *et al. Doklady AN SSSR* **306**, 844 (1989).
- Moroso, L. L., Kuzmin, V. V. & Goldanskii, V. I. *Doklady-Biophys. (Proc. Acad. Sci. USSR)* **274**, 55 (1984); **275**, 71 (1984).
- Zahle, K. & Grinspoon, D. *Nature* **348**, 157 (1990).
- Goldanskii, V. I. *et al. Doklady-Biochemistry (Proc. Acad. Sci. USSR)* **207**, 218 (1972).

DNA strand asymmetry

SIR — We reported¹ an apparent inequality in the mutation rates of the two strands of DNA. Our analysis was based on seven DNA sequences in an intergenic region within the β -globin gene complex of six primate species. Our observation was in good agreement with an independent study of the $\phi\eta$ region, 4 kilobases away². Since then, the DNA spanning a 11.5-kb region that encompasses the two subregions has been sequenced; analysis of the whole region³ does not show the trend observed in the two subregions^{1,2}. We have partitioned this data into four subregions for more detailed analysis.

The results show that the inequality in ref. 1 is restricted to the last 3 kb. The other three subregions all show weak asymmetry in the opposite direction (C.-I. W. and T. Baldanzi, unpublished results). One explanation is that the locations of replication origins, at least in the germ cells, have not been conserved since

the divergence of these species. As a consequence, regions of homologous DNA strands may be replicated as a leading strand in one species and as a lagging strand in another species, obscuring the asymmetry reported in ref. 1.

This possibility becomes greater when a longer sequence is examined. Umek *et al.*⁴ have reviewed cases of nonspecific and stochastic usage of replication origins in higher eukaryotes. It is also possible that mammalian male germ cells may recruit many secondary origins for rapid replication. For sequence analysis, it is therefore most promising to compare molecules with known fixed origins, such as bacterial chromosomes or mitochondrial DNA.

CHUNG-I WU

Department of Biology,
University of Rochester,
Rochester, New York 14627, USA

- Wu, C.-I. & Maeda, N. *Nature* **327**, 169–170 (1987).
- Koop, B. F., Goodman, M., Xu, P., Chan, K. & Slighton, J. L. *Nature* **319**, 234–238 (1986).
- Goodman, M., Koop, B. F., Czelusniak, J., Fitch, D. H. A. & Slighton, J. L. *Genome* **31**, 316–335 (1989).
- Umek, R. M., Linskens, M. H. K., Lowalski, D. & Huberman, J. A. *Biochem. biophys. Acta* **1007**, 1–14 (1989).

AZT before AIDS

SIR — Our early work on azidothymidine (AZT) provided the first demonstration of its antiretroviral activity and indeed constituted the only research of the time to suggest what would later turn out to be efficacy at treating people with AIDS.

AZT was born via chemical synthesis by J. Horwitz *et al.* (*J. org. Chem.* **19**, 2076; 1964). After a relatively uneventful childhood, its biological potential was first clearly recognized in the early 1970s and by its tenth birthday firmly established (W. Ostertag *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4980; 1974). In those experiments, we examined the effect of AZT on the replication of a murine retrovirus complex consisting of spleen focus-forming and leukaemia viruses and made the crucial observation that AZT interfered with virus replication quite dramatically but had very little hindrance of cell growth even at relatively high concentrations of 2.5 $\times$ 10⁻⁴ M.

These observations, showing AZT's high toxicity for the virus but low toxicity for the cell, suggested for it a career of high promise. Indeed, on its twenty-first birthday it came of age when H. Mitsuya *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **82**, 7096; 1985) demonstrated its usefulness against human T-lymphotropic virus type III. It is gratifying that the effects of AZT in a murine retrovirus system have faithfully reproduced with a human retrovirus and, once again, mouse has proved its worth to man.

S. K. DUBE

University of Maryland,
College Park, Maryland 20742, USA

W. OSTERTAG

Heinrich Pette Institute,
2000 Hamburg 20, Germany

Calcium control and InsP_4

SIR — In their paper on page 162 of this issue¹, Bird *et al.* claim to demonstrate that inositol trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$) alone can account for calcium-ion entry in stimulated cells, with no need for a contribution from the tetrakisphosphate $\text{Ins}(1,3,4,5)\text{P}_4$. But the reported experiments do not actually address the issue of the physiological role of $\text{Ins}(1,3,4,5)\text{P}_4$ *in vivo*. What the authors do show is that introducing high doses of $\text{Ins}(1,4,5)\text{P}_3$ or $\text{Ins}(2,4,5)\text{P}_3$ as a bolus into cells can cause calcium-ion entry, an observation that has been made by others in several experimental systems (see ref. 2 for review).

Under another experimental paradigm very similar to that of Bird *et al.*, $\text{Ins}(1,4,5)\text{P}_3$ did not cause calcium-ion entry, but needed help from $\text{Ins}(1,3,4,5)\text{P}_4$ (ref. 3). This phenomenon may, of course, be an artefact (an explanation implicit in the paper by Bird *et al.*), but if this is so, is it likely to be a quantitative distortion of an existing mechanism or a new phenomenon created *de novo* by experimental conditions?

Credulity in the latter alternative is stretched to the limit by the observation that the clear requirement for $\text{Ins}(1,3,4,5)\text{P}_4$ could not be substituted by $\text{Ins}(1,4,5)\text{P}_3$, $\text{Ins}(1,3,4)\text{P}_3$, $\text{Ins}(2,4,5)\text{P}_3$, InsP_3 , or $\text{Ins}(1,3,4,5,6)\text{P}_5$, all at concentrations five- or more-fold than the fully effective dose of $\text{Ins}(1,3,4,5)\text{P}_4$ (ref. 4), a pharmacology which matches an InsP_4 -binding protein found in many tissues⁵. These data show

clearly that $\text{Ins}(1,3,4,5)\text{P}_4$ can control calcium entry, and suggest that if this phenomenon is an artefact, it is one of degree only. Moreover, if, as Bird *et al.* claim, $(1,4,5)\text{P}_3$ is responsible for everything in an intact cell *in vivo*, where is the selective pressure for $\text{Ins}(1,3,4,5)\text{P}_4$ (and its receptor) to evolve and remain in existence for hundreds of millions of years?

If experimental artefacts are indeed distorting the picture, I suggest that there is a more plausible alternative for their nature in the possibility that the receptors for $\text{Ins}(1,4,5)\text{P}_3$ and $\text{Ins}(1,3,4,5)\text{P}_4$ interact, and their dissociation causes calcium entry^{2,6}. *In vivo*, this dissociation would be promoted by the ligands (InsP_3 and InsP_4), but any experimental manoeuvre that separates these two proteins will cause calcium entry irrespective of the presence of $\text{Ins}(1,3,4,5)\text{P}_4$. Thus one need propose only that various experimental regimes influence quantitatively the interaction between two allosteric proteins to reconcile the apparent contradictions.

R. F. IRVINE

AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK

1. Bird, G. S. J., Rossier, M. F., Hughes, A. R., Shears, S. B., Armstrong, D. L. & Putney, J. W. Jr *Nature* **352**, 162–165, (1991).
2. Irvine, R. F. *Bioessays* (in the press).
3. Morris, A. P., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *Nature* **330**, 653–655 (1987).
4. Changya, L., Gallacher, D. V., Irvine, R. F., Potter, B. V. L. & Petersen, O. H. *J. Membr. Biol.* **109**, 85–93 (1989).
5. Theibert, A. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3165–3169 (1991).
6. Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1991).

Induced molecule self-organization

SIR — In his comment¹ in News and Views on our work on the encapsulation of anionic guests in inorganic host molecules, Mitchell wondered whether the vanadium oxide host shells we synthesized exist without the encapsulated species. In answer, we have no evidence that they do. We find the encapsulation is due to a new type of self-organization process leading to unusual weak repulsive interactions in molecular species and to novel topological structures.

With the self-assembly process, almost any hollow sphere (cluster shell) can be generated, depending on size, shape and charge of the encapsulated anionic species X (the template), by using square-pyramids (of OVO_4 ; Fig. 1) as the repeating building block. The central anion is responsible for the architecture of the system, determining the number of linked units as well as their type of linking. A new and typical example is $[\text{H}(\text{VO})_{18}\text{O}_{26}(\text{NO}_3)]^{10-}$ (Fig. 2) with an ellipsoidal structure, which has NO_3^- as the encapsulated ion.

Characteristic of this type of species is the rather large separation between negatively charged centres (here O...O) due to very weak (repulsive) forces, never before

reported. The weak repulsive interactions can also allow unusually high 'coordination numbers' as high as 24 (see below); the highest known conventional coordination number is 12². These types of weak interactions are probably important in the poorly understood mechanism of anion exchange through membranes which occurs very rapidly.

Interesting topological considerations should also be mentioned. By linking 18

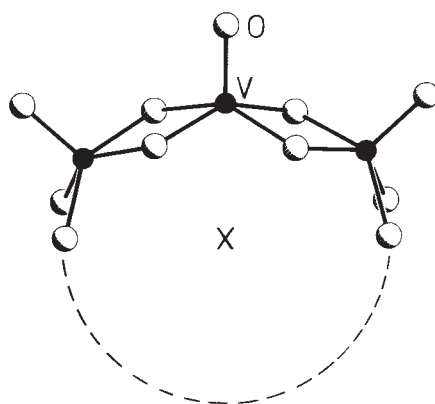


FIG. 1 Building molecular shells from OVO_4 tetrahedra.

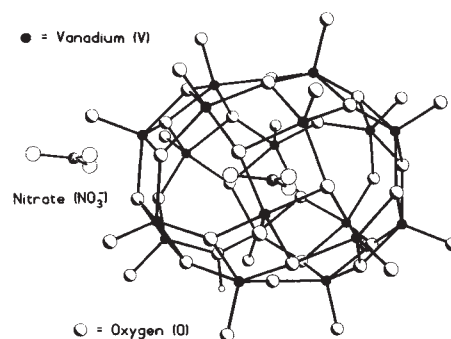


FIG. 2 The new inclusion species $[\text{H}(\text{VO})_{18}\text{O}_{26}(\text{NO}_3)]^{10-}$, prepared by R. Röhling in our group. Notable in the structure is the large distance between the O atoms in shell and in the central anion, 280 picometres. With a more highly charged anion, such as CO_3^{2-} , the interaction with the V^{5+} centres in the shell is increased and the separation becomes less.

tetragonal OVO_4 pyramids, spherical hollow spheres, $[(\text{VO})_{18}\text{O}_{24}(\text{X})]^{n-}$, spanned by 24 O atoms can be generated³. Depending on the anion X the O atoms of the cluster shell with the same stoichiometry can either adapt the form of one of the 13 archimedean solids (the rhombicuboctahedron with 18 squares and 8 triangles) or the mysterious so-called fourteenth archimedean solid (or pseudo-rhombicuboctahedron). These cluster shells can be transformed into each other by rotating the upper hemisphere of the cluster by 45°.

ACHIM MÜLLER

Faculty of Chemistry,
University of Bielefeld,
PO Box 86 40,
D-4800 Bielefeld 1,
Germany

1. Mitchell, P. C. H. *Nature* **348**, 15–16 (1990).
2. Wells, A. F. *Structural Inorganic Chemistry* (Clarendon, Oxford, 1989).
3. Müller, A., Penk, M., Röhling, R., Krickemeyer, E. & Döring, J. *Angew. Chem. Int. Edn. Eng.* **29**, 926–927 (1990).

SIR — Although we were flattered to have our research cited in Mitchell's News and Views article¹ on new guest-host chemistry, the article may have left the mistaken impression that purely inorganic analogues of organic crowns and cryptands were hitherto unknown.

Inorganic crowns and cryptands have been investigated for at least 15 years. Interest was sparked initially by the biological activity of the sodium cryptate $[\text{NaW}_{21}\text{Sb}_9\text{O}_{86}]^{18-}$ (ref. 2), a species that has subsequently achieved notoriety as HPA23 in AIDS therapy. Larger, crown-shaped cryptands have been reported more recently: $[\text{P}_5\text{W}_{30}\text{O}_{110}]^{15-}$ (ref. 3); $[\text{As}_4\text{W}_{40}\text{O}_{140}]^{28-}$ (ref. 4) and $[\text{P}_8\text{W}_{48}\text{O}_{184}]^{40-}$ (ref. 5). The last two species have multiple cation binding sites in the interiors of their crown structures; the $[\text{As}_4\text{W}_{40}\text{O}_{140}]^{28-}$ cryptand even displays allosteric binding effects.

Inorganic complexes containing encapsulated anions were first characterized by Keggin⁶ in 1933. Today, the number of inor-

ganic 'guest-host' complexes known to contain anions is vast⁷. Unlike the cation crown complexes just described, most of these anion complexes contain nonexchangeable ions entrapped within polyhedral cages. The $\text{PMo}_{12}\text{O}_{40}^{3-}$ ion, for example, has been described⁸ as a clathrated PO_4^{3-} ion. A new type of anion complex was recently prepared, however, in which anions like Cl^- and NO_3^- are bonded to shallow basket-like inorganic frameworks⁹. These may turn out to be true guest-host as opposed to host-guest complexes.

VICTOR W. DAY

Department of Chemistry,
University of Nebraska,
Lincoln,
Nebraska 68588, USA

WALTER G. KLEMPERER

Department of Chemistry,
University of Illinois,
Urbana,
Illinois 61801, USA

OMAR M. YAGHI

Department of Chemistry,
Harvard University,
Cambridge,
Massachusetts 02138, USA

MITCHELL REPLIES — I omitted heteropolys from my article on anions as guests in inorganic hosts because of space limitations and because the heteropolys are so well known. Nevertheless, my reference to the paper by Day *et al.*¹⁰ provided a lead-in, although I ought to have added Michael Pope's excellent monograph⁷.

The particular interest of Müller's work, and my concern, was the encapsulation of guest anions which were weakly bound and entities chemically distinct from the host framework. Whether a particular anion is a guest, or part of the overall structure, arises with a number of heteropolys. For the phosphomolybdates and tungstates, $[\text{PM}_{12}\text{O}_{40}]^{3-}$, the question is whether the oxygens of the phosphate are part of the framework — that is, whether the phosphate anion is a distinct entity. Clark and Hall, who solved the structure¹¹, do indeed refer to the phosphate as a guest species for $\text{M} = \text{Mo}$ but not for $\text{M} = \text{W}$. With regard to 'anion complexes', Heinrich *et al.*⁹ did not use the term 'guest-host', referring to their compounds as aggregates.

Let us agree about the questions to be asked. Which came first, the host or the guest (is the guest a template)? How far is the encapsulation of the anionic guest dependent on a structural match with the host cavity (molecular recognition)? Is the host stable without the guest? And finally, particularly important for applications, can we move the guest out of the host cavity (or is it really a 'hostage')? My article and this correspondence have provided a perspective on inor-

ganic guest-host complexes encompassing a number of apparently disparate structural types and some familiar compounds such as the heteropolys which had not been thought of in this way. We can now look forward to developments and applications.

PHILIP C. H. MITCHELL

Department of Chemistry,
The University,
Whiteknights,
Reading RG6 2AD, UK

High salinity in the North Sea

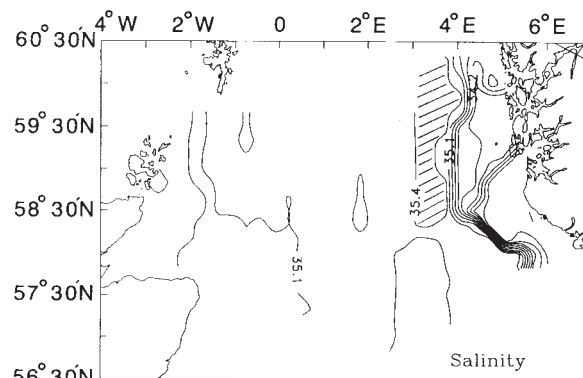
SIR — The salinity of sea water in the north-east Atlantic has been documented since 1902; several large-scale fluctuations have been detected. During the mid 1970s the salinity declined to low levels not encountered since 1908 (ref. 1), increasing to more usual levels during the 1980s. Atlantic water flows into the northern North Sea between the Orkney and Shetland isles, and between Shetland and Norway, and a corresponding outflow from the North Sea occurs along with Norwegian coast. Changes in salinity in the northern North Sea during the 1970s paralleled those in the North Atlantic², and were correlated with changes in fish stocks and plankton composition^{3,4}. We now report the discovery in January 1990 of an inflow of water of exceptionally high salinity into the northern North Sea.

We conducted a survey in the northern North Sea from the vessel *British Enterprise III*. Salinity and temperature at a depth of 5 m were recorded from sensors towed alongside the ship, and water currents measured with a multiple beam echo-sounding device (acoustic doppler current profiler). The highest recorded salinities (greater than 35.45, maximum value 35.471) were found in a strip approximately 20-km wide parallel to the Norwegian coast, corresponding to the area of inflow from the Atlantic (see figure). The water currents in the high salinity region were southerly, whereas 10–20 km to the east in the lower salinity (33.5–35.0) Norwegian coastal water, the flow was northerly.

The maximum salinity was higher than any value measured in the North Sea by the Marine Laboratory, Aberdeen, since records began in 1920 (previously recorded maximum: 35.390 in August 1968). Salinities greater than 35.40 have been recorded on a number of occasions during this period in the Atlantic between Shetland and the Faroe Isles and at the edge of the continental shelf off the northwest of Scotland (in 1929, 1939, 1959, 1969, 1971, 1983–85 and 1989), but none has exceeded 35.47. It therefore seems

that the salinity of the water entering the North Sea in January 1990 was perhaps the highest this century.

The late 1980s were characterized by an increasing frequency of westerly weather patterns in the North Atlantic following a minimum in the 1970s (ref. 4). It is possible that the high salinity of the water entering the North Sea in 1990 signals changes in current patterns in the Atlantic Ocean related to the changing weather conditions, bringing water from more southerly latitudes into the north-eastern Atlantic. Abundance of plankton in the northern North Sea decreased during the



Distribution of salinity at 5-m depth in the northern North Sea in January 1990 drawn from data collected at 180-s intervals throughout the survey. Shaded area, water of salinity greater than 35.40.

1970s and has increased since 1980 (ref. 5), in parallel with the increasing incidence of westerly weather⁴. There have also been latitudinal fluctuations in the spawning of herring and sprat³ and changes in the breeding success of some sea birds over similar time-scales⁴, suggesting a link between North Sea ecology and climate changes in the North Atlantic. The exceptionally high salinity in 1990 may signal the start of further changes.

M. R. HEATH

E. W. HENDERSON

G. SLESSER

Marine Laboratory, PO Box 101,
Aberdeen AB9 9DB, UK

E. M. S. WOODWARD

Plymouth Marine Laboratory,
Prospect Place, Plymouth PL1 3DH, UK

- Mitchell, P. C. H. *Nature* **348**, 15–16 (1990).
- Fischer, J., Ricard, L. & Weiss, R. *J. Am. Chem. Soc.* **98**, 3050–3052 (1976).
- Alizadeh, M. H., Harmalkar, S. P., Jeannin, Y., Martin-Frère, J. & Pope M. T. *J. Am. Chem. Soc.* **107**, 2662–2669 (1985).
- Robert, F., Leyrie, M., Hervé, G., Tézé, A. & Jeannin, Y. *Inorg. Chem.* **19**, 1746–1752 (1980).
- Content, R. & Tézé, A. *Inorg. Chem.* **24**, 4610–4614 (1985).
- Keggin, J. F. *Nature* **131**, 908–909 (1933).
- Pope, M. T. *Heteropoly and Isopoly Oxometalates* (Springer, Berlin, 1983).
- Boeyens, J. C. A., McDougall, G. J. & Smit, J. V. R. *J. Solid State Chem.* **18**, 191–199 (1976).
- Heinrich, D. D., Folting, K., Streib, W. E., Huffman, J. C. & Christou, G. *J. Chem. Soc., Chem. Commun.* 1411–1413 (1989).
- Day, V. W., Klemperer, W. G. & Yaghi, O. M. *J. Am. Chem. Soc.* **111**, 5959–5961 (1989).
- Clark, C. J. & Hall, D. *Acta Cryst.* **B32**, 1545–1547 (1976).

- Dooley, H. D., Martin, J. H. A. & Ellett, D. J. *Rapp. P.-v. Reun. Cons. Int. Explor. Mer* **185**, 179–187 (1984).
- Dickson, R. R., Jones, S. R. & Lee, A. J. *ICESCM 1984/Gen. B* (1984).
- Corten, A. *Neth. J. Sea Res.* **25**, 227–235 (1990).
- Aebischer, N. J., Coulson, J. C. & Colebrook, J. M. *Nature* **347**, 753–755 (1990).
- Colebrook, J. M. *Hydrobiologia* **142**, 309–325 (1986).

Exceptional treasures

Michael Neve

Bully for Brontosaurus. By Stephen Jay Gould. *Radius/Norton*: 1991. Pp. 540. £16.99, \$22.95.

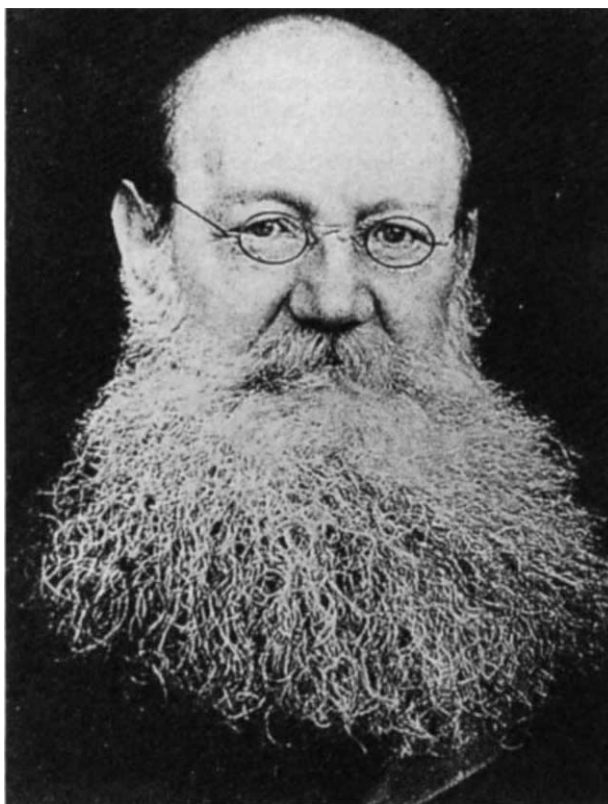
'TREASURE your exceptions'. This thought, which appears in the middle of one of the collected essays that make up Stephen Jay Gould's latest volume, might be said to represent the heart of all his arguments. *Bully for Brontosaurus* displays Gould's philosophy and enthusiasms to the full, more so than earlier volumes (some of which he now wishes would go away) and with more variety and less implausibility than his successful *Wonderful Life* (Norton, New York; for review see *Nature* 342, 303; 1989).

With this book Gould stands up and, in many ways, stands alone. He performs his usual admirable service of passing on views from the academy into the wider world (the prose is close to vulgarity in the most honourable sense); he has also taken his arguments about evolution and about the history of science to be internally coherent, and has put his faith in them. At various points his writing reaches for remarkable statements of affirmation, of warnings against hubris, of how to endure life's distresses and glories. Some of the language — deployed of course with irony but also with feeling — is biblical. All of it insists that we benefit permanently from pondering the nature of evolution.

Gould's account of evolution and of how to see it without illusion is now fully formed. The great mistake in his view is to be insufficiently interested in 'mistakes'. What looks primitive, ill-arranged, out of place or clumsy, is for Gould the very place where the evolutionist must look for the evidence. If we persist in this task and begin to break through the fantasy of anthropocentric ascendancy that is the opposite of Gould's philosophy, a new way of seeing is possible — we get to this new place, not by great schemes of progress, or by writing off various groups of organic beings as 'lower' or 'maladapted', but by grubbing around in the practical task of seeing how nature makes and cobbles things together. And how this is quite wonderful. Gould once again shows what kind of darwinian he is: one less impressed by "large and portentous movements", based on crude adaptationism, but instead a student of "tiny quirks and circumstances that appear insignificant at the time but cascade into later, and unpredictable, prominence". Readers of *Nature* will know that this position of Gould's is not

without difficulties, some of them major. But the persistence with which he has followed his star is impressive and undeniable.

As the (for once) well-informed blurb on this book makes clear, its central theme is the interplay between contingency and necessity in evolution. There are many essays that speak to this, and the best of them seem to benefit from a visit Gould made to those homes of natural magnificence (and pathos), Australia and New Zealand. The platypus,



Russian revolutionary anarchist Petr Kropotkin — neither crackpot nor loser.

to quote from a famous English historian writing in a very different context, is rescued "from the enormous condescension of posterity"; the glow-worms of New Zealand are re-examined; and the frogs of Australia become a model "for the introduction of creativity and new directions in evolution". The essay on Australian frogs also carries a sad postscript that reminds the reader of one of Gould's persistent themes: for unexplained reasons, amphibian populations are disappearing throughout the world. There is also a delightful account of the size of kiwi's eggs in New Zealand, "this easy land of no natural predators, where a female might waddle without fear as an enormous egg distends her abdomen during passage down the

oviduct". (The kiwis themselves had become smaller, while their eggs remained their former size.)

Gould's concern with the quirks of adaptation as against the 'adaptive' selecting out of quirks, shades into his concern about myths in the history of science. Generously bringing to attention the work of historians of science, he pays especial attention to this here. Historical myths are like false forms of evolutionary adaptation — they conceal the confusions and the tangled bank of anecdote and reportage that actually make up the story. Kropotkin, the Russian anarchist, naturalist and author of *Mutual Aid* (McClure Phillips, New York, 1902), was neither a crackpot nor a 'loser' in the history of evolutionary debates. And at the famous encounter in Oxford in 1860, T. H. Huxley was in no sense a 'winner' in the debate with

the Bishop of Oxford, Samuel Wilberforce. More truthfully, we hardly know what actually was said. Instead, we can investigate the story about the stories that were made up to prove other points. Even more impressively (and this is an excellent piece of his own work), Gould writes about evolution's enemy William Jennings Bryan, populist and combatant in the Tennessee 'monkey trial' of 1925. Gould brings Bryan to life, describing him as a complex figure who was protective of his people, dismayed that they would be miseducated, and yet blind to the evolutionary story.

Gould is using a form of historical double-header, both looking at natural oddities as points of growth, not culs-de-sac, and at historical ideas of victory as complex forms of mythology. He does so with the enthusiasm of the baseball fan (it is no secret that he regards Joe DiMaggio as "the model and hero of my life", the man who took baseball, for a moment, outside itself).

The zest and the self-generated pleasure that Gould brings to his various tasks cannot dodge some

real problems. It looks very possible that the fossil arthropods of the Burgess Shale (described in *Wonderful Life*) may not be as different, as remote and strange, as Gould proposed (see *Nature* 351, 184 and 225; 1991). Gould has exaggerated their strangeness, his critics would say, because he has his own fantasy of evolution which is projected onto the deep past. In the English scientific context, as described by John Maynard Smith among others, the complexities of conventional adaptationism may be underestimated by Gould, leaving his arguments for historical 'contingency' somewhat light-headed, a quirk, not of natural forms, but of his own. Gould exaggerates the contingent to develop a natural philosophy that only seems

different. Close up it is less so. And for environmentalists of a certain persuasion, Gould can seem eerily indifferent to the language of crisis and denudation, and to the cries of exhausted nature and scared humanity. Nature's may be a darker story than Gould thinks, more 'cruelly' darwinian.

Bully for Brontosaurus is the book that answers these doubts. In the face of many distresses (not least his own), Gould has allowed himself a good old-fashioned response to the world he sees, not least in the achievement of *Voyager* in photographing Neptune. He feels wonder. It may be an irony

that having often reminded us that we make up our own values, independent of nature (which will survive us, having long lived without us), he should allow himself such a fulsome reaction as that of permanent wonder. But these fine essays show that he entirely deserves so to do. Treasure your exceptions. □

Michael Neve is in the Department of Anatomy and Developmental Biology, University College, Gower Street, London WC1E 6BT and at the Wellcome Institute for History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Weighing up the costs

Darryl T. Gwynne

The Evolution of Parental Care. By T. H. Clutton-Brock. Princeton University Press. 1991. Pp.352. Hbk \$49.50, pbk \$19.95.

A MALE stickleback guarding eggs in a nest; a dying female mite exhibiting an extreme form of 'live birth' as she is devoured by her own progeny from the inside; and a male blister beetle inseminating his mate with Spanish fly, a chemical transferred to and conferring protection on eggs. Such diverse forms of reproductive effort all represent parental care in a broad and descriptive sense. From this concept of parenting, common themes can emerge, in particular the similar influence of ecological variables on the parental aspects of life histories.

Studies of parental expenditure may deal with its adaptive significance or its effect on sexual competition (mating systems). Clutton-Brock is almost exclusively concerned with the former topic and outlines five main areas for examination: the costs and benefits of parental care; interspecific variation in the extent of care; variation in the sex assuming the role of direct care of the progeny; how care is adjusted in its benefits to offspring and cost to parents; and how parents divide investment between sons and daughters. But the scope of his book is much broader than these areas may indicate and includes topics as disparate as egg size and interspecific variation in the composition of mammalian milk. The book is a good reference for the diverse forms of parental expenditure in animals, although the short subject index tends to limit this function. Many groups are considered and a quick glance through the excellent illustrations by Dafila Scott gives a flavour of the diversity of taxa covered. An obvious omission is plants; with such a broad concept of parental care, botanical information would have been useful to topics such as female gamete size and sex allocation.

Clutton-Brock discusses both problems encountered in answering certain questions

and where further investigation is needed. But more could have been done to point out how particular past studies have answered questions. An example concerns a question Clutton-Brock considers among the most fascinating: what factors determine the extent to which males and females are involved in direct (behavioural) care of progeny? One area of interest stems from the fact that in most animal groups, exclusive male care tends to occur when there is external fertilization of eggs. Several hypotheses have been advanced to explain this association; the fishes represent a key testing ground because of the diversity of fertilization modes and direct parental-care states within this group. Although Clutton-Brock outlines these hypotheses, he does not point

at least one other explanation, one that involves the influence of social environment. Models of the evolution of female mating preferences predict that populations with similar ecologies can evolve very different forms of male secondary sexual traits. A history of female preference for nurturant males in a population should perhaps be considered as a possible cause of sexual dimorphism in parental-care roles. It is known that male care can attract mates; male sticklebacks with eggs in the nest are preferred by females. Such males are even known to steal eggs from other nests, an act reminiscent of male bower birds stealing 'secondary sexual' attractive ornaments from the bowers of rivals.

One of the benefits of a review of a large subject is the clarification of questions for future research. In his concluding chapter, Clutton-Brock points to several such areas, including the need for improved measures of the costs and benefits of parental care and for certain theoretical analyses. Investigation of these areas would presumably involve the use of experimental and modelling methodologies. In addition to these, there is a need for phylogenetically controlled comparative approaches which would be particularly useful in understanding the diversity of parental-care strategies in animals. Such approaches rely on the reconstruction of phylogenies to determine both evolutionary

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sticking around — why do male sticklebacks guard the eggs in the nest?

out that published studies report (statistical) separation not only of certain hypotheses that explain the association between care states and modes of fertilization but also of hypotheses that predict the direction of change of these states in evolutionary time.

As Clutton-Brock points out, exceptions to these trends, such as cases of exclusive male care in internally fertilizing species, also need explaining. He suggests that paternal care is retained in such cases because of ancestry or ecological pressures such as high costs of egg production that emancipate females from direct-care duties. But there is

changes in parental characters and the associations of these characters with other variables. It is this aspect, the extent to which further research is stimulated, that is perhaps the best measure of the success of a book such as this. This volume generates many avenues for future study and thus should be required reading for all behavioural and evolutionary ecologists. □

Darryl T. Gwynne is in the Zoology Department, Erindale College, University of Toronto, Mississauga, Ontario L5L 1C6, Canada.

Bright spark of the electric age

Willem Hackmann

Ferranti and the British Electrical Industry, 1864–1940. By J. F. Wilson. Manchester University Press: 1991. Pp.165. £12.95. Published by St Martins Press in the United States, \$24.95.

WHEN Sebastian Ziani de Ferranti died on 13 January 1930, at the age of 65, the British electrical industry lost one of its most respected figures. He combined his family's interest in art (his father settled in Liverpool as a pioneer photographer) with a childhood fascination for technology. His parents fostered this interest with books like Samuel Smiles' *Life of the Stephensons*.

Indeed, Ferranti's own life seems straight from the pages of one of Smiles' popular Victorian biographies on notable engineers. His teenage bible was Ganot's highly readable *Treatise on Physics*, which was full of the technological marvels of the day. It also contained one of the most detailed descriptions of the theory of electromagnetism available in the 1870s, totally misunderstood by the young Ferranti, as he wrote to his parents that he was in the process of inventing a perpetual motion machine. His lack of academic application (as shown by his quaint phonetic spelling) must have been something of a headache to the teachers of his Catholic school, St. Augustine's at Ramsgate, but the institution was liberal enough to allow him to follow his technical bent. (When I went to the same school in the mid-1950s, Ferranti had become a shadowy memory. Like him, I was given the key to the well-equipped laboratory — much of it from Ferranti's days — so that I could follow my own interests, but I am afraid that is where the similarity ends.) Ferranti was imbued strongly with the belief of his age that science-based technology would be a 'blessing to the world'. He had supreme confidence both in himself (a very Victorian virtue) and in the role electricity would play in the modern world. Throughout his career as an electrical manufacturer, virtually from the birth of the industry in 1882 to the implementation of the 1926 Electricity (Supply) Act, he pursued the goal of the 'all electric' age.

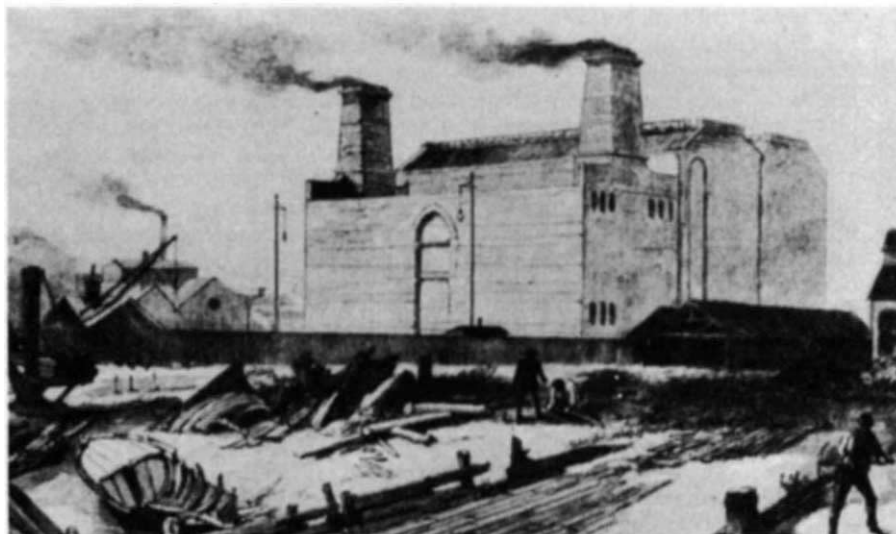
What makes Wilson's book different from previous biographies is that it sets Ferranti's career against the economic background. Indeed, this is as much a business history as a biography, and closely follows the intention of the series of which this volume is a part, to illuminate the role of businessmen and business institutions in forming the modern social fabric. Ferranti was deeply involved both in the development of an efficient supply industry and in the establishment of a viable electrical-engineering industry. These

motivating forces were clearly illustrated in all he did, from his construction of the world's first modern power station at Deptford in the late 1880s to the emergence of Ferranti Limited as one of the most progressive British electrical firms.

One of Wilson's aims was to examine, in the light of Ferranti's career, the common stereotype of the Victorian and Edwardian entrepreneur as being generally ignorant of matters scientific (because of a supposedly classical public-school education), failing to exploit products based on the new technologies, and being unwilling to invest in the latest efficient mass-production techniques. Ferranti certainly exhibits none of these faults, which allegedly started Britain's decline as a world economic power. As an engineer he was innovative, forward-looking and quick to exploit (and improve on) good

having small local stations supplying low-voltage direct current. Instead, Ferranti argued that the country should be served by a system of large stations with easy access for their fuel (coal), supplying alternating current at very high pressure which was stepped down at local stations.

The 'Battle of the Systems' was ultimately resolved in Ferranti's favour with the formation of the National Grid, but at the time Deptford became a very expensive failure because of government legislation in favour of local supply systems. The technical problems were enormous. For instance, no tools were big enough to construct the huge alternators so these had to be constructed in house. The lathe required to turn the main shaft was of the same dimension as that used for turning a 100-ton gun at Woolwich. In May 1891, the project to construct the giant



Sketch of the Deptford power station from the *Illustrated London News*.

ideas. His 200 or so patents on such diverse mechanisms as alternators, transformers, arc lamps, electrical supply meters, high-voltage cables, fuses, switchgear, an electrical furnace, a high-temperature steam turbine, stop valves and a high-speed cotton spinning frame, testify to this. Not all these inventions bore fruit, for a variety of reasons. As a businessman he was conservative in financial and administrative matters and did not accept advice easily. This tendency was in part responsible for the series of financial crises that affected his business dealings, and which are well described in this book. It also constantly weakened his ability to stand in competition with such giants as the German AEG or the American General Electric Company, and brought him, in 1903, to the brink of ruin. In this respect he was perhaps also typical of his generation.

One of the most endearing aspects of his character was his sheer enthusiasm, which he could communicate to others. Who else could have persuaded hard-headed businessmen to invest in the construction of the largest power station in the world at that time. Deptford was 40 years before its time and ran counter to the accepted wisdom of

10,000-volts, 10,000 horse-power steam engines and alternators was abandoned. Had they been completed, each alternator would have weighed nearly 500 tons, with the armature and shaft accounting for 225 tons. Edison was impressed when he visited the site in 1889, and congratulated the 'Michaelangelo' of Deptford but disagreed with the project.

Wilson concludes that Ferranti's problem lay in matching his ideas to an unresponsive environment. He was not a commercial engineer and his company only began to exploit his products successfully after chartered accountants joined the board. Nor was he good with labour relations (an aspect of his character that could have been developed further), but he did work hard for standardization in the electrical industry. His work has certainly touched all our lives, not least the small air turbines developed for cotton spinning, which later found application in dentists' drills. □

Willem Hackmann is at the Museum of the History of Science, University of Oxford, Old Ashmolean Building, Broad Street, Oxford OX1 3AZ, UK.

Raking over the coals

Lee Loevinger

Science, Ideology, and the Media: The Cyril Burt Scandal. By Ronald Fletcher. Transaction Publishers: 1991. Pp. 419. \$29.95, £21.95.

A LEGEND of modern science is the Cyril Burt scandal — an eminent scientist posthumously charged with fabricating data. Ronald Fletcher now says the charge is untrue: a spurious product of politics and publicity intruding into science.

Burt (1883–1971) was a pioneer in quantifying psychology, studying individual differences and mental testing. Knighted and the recipient of many other honours during his lifetime, he was described in a *Times* obituary as “Britain’s most eminent educational psychologist”. He also believed that intelligence was largely, although not entirely, determined by heredity, a position supported by several papers reporting kinship correlations of intelligence, the most significant of which involved monozygotic twins separated at birth or soon thereafter and raised in different environments.

Strongly opposed to Burt’s position were those who contended that mental abilities were environmentally determined, that IQ tests and other hereditary data were unreliable, and that there really was no single trait identifiable as intelligence. The most vehement environmentalists were left-wing egalitarians who accused Burt of justifying class distinctions.

The first published criticism of Burt’s kinship correlations was by Arthur Jensen, who noted 20 pairs of invariant correlations for samples of varying sizes in Burt’s publications. In 1974, Leon Kamin wrote a slim volume entitled *The Science and Politics of IQ* (Erlbaum) in which he attacked Burt’s data, arguing that IQs are never free of policy implications and ideology. Criticism of Burt erupted into scandal on 24 October 1976 when Oliver Gillie, medical correspondent of the *Sunday Times*, published an article beginning: “The most sensational charge of scientific fraud this century is being levelled

against the late Sir Cyril Burt”. Gillie charged that Burt’s correlations were based on fabricated data and that two women Burt claimed as assistants in gathering data never existed.

The controversy simmered until 1979 when L. S. Hearnshaw published a biography of Burt (*Cyril Burt: Psychologist*, Cornell University Press, 1979) in which he observed that Burt was “anathema to left-wing egalitarian critics” and that Kamin was biased and captious, but found Burt guilty of fabricating data on monozygotic twins. Hearnshaw credited Burt with being an imaginative pioneer who contributed to psychology but said that his failings were due to a “marginally paranoid condition”. The Hearnshaw biography established the fraud charge for all journalists and most scientists. Thus the legend began.

In 1989, Robert Joynson, in a book entitled *The Burt Affair* (Routledge, 1989) concluded that Hearnshaw’s sources did not support the charges and that Burt’s work was more credible than that of his critics (for review see *Nature* 340, 439; 1989). In *Science, Ideology, and the Media*, Ronald Fletcher presents an even more exhaustive array of evidence to conclude that Burt’s only deception was publishing papers he wrote under the name of an assistant. Fletcher is deliberately legalistic, writing as counsel for the defence with cross-examination questions for each of Burt’s critics. This approach is a dubious literary device, but Fletcher is right in treating the controversy as similar to common litigation: his marshalling of evidence is sufficient to persuade a lawyer with substantial trial experience and a degree in psychology that his conclusions are correct.

It is apparent that the fraud charge rests entirely on circumstantial evidence, mainly the great improbability of invariant correlations being found in a series of enlarging samples. But to draw a valid inference from circumstantial evidence, it is necessary to consider all the circumstances, and Fletcher presents many that Burt adversaries have ignored.

The invariant data appear in papers published up to 1966, five years before Burt’s death, but the fraud charges were not made until five years after his death, and then by a journalist seeking a sensational story. The charge of nonexistence of lady assistants, featured in Gillie’s story, is disproved both by the testimony of those who knew the women and by written records. Cartons containing some of Burt’s accumulated data, which might have provided direct evidence concerning assailed correlations, were destroyed shortly after his death by Burt’s secretary on the advice of Liam Hudson, one of Burt’s vehement adversaries.

The crux of Kamin’s attack, underlying Gillie’s story, was that in 1943 Burt reported 15 pairs of separated monozygotic twins with an IQ correlation of 0.77. In a 1955 paper, the sample size was 21 and the correlation

was 0.771. In a 1958 paper, the sample size had risen to ‘over 30’ and the correlation was 0.771; and in the 1966 paper the sample size was 53 and the correlation was again 0.771. But Kamin himself was careless. There was no separate 1958 report of data. The 1958 paper was the publication of a lecture by Burt delivered in 1957 in which he reproduced the entire table of figures published in 1955, and then went on to say that by 1957 he had additional cases of monozygotic twins bringing the total to over 30.

On the basis of a detailed analysis, Fletcher finds the only suspicious invariant correlation to be the repetition in 1966 of 0.771 first found in 1955. But in 1966, Burt was 82 years old, with declining powers of concentration, doing all his calculations by hand, and writing in haste to reply to other articles. Because the increasing sample sizes were cumulative additions, it is more plausible to suppose that Burt added additional cases to his collection without bothering to calculate new correlations than it is to conjecture that the longtime editor of the *British Journal of Statistical Psychology* would stupidly fabricate improbably invariant correlations in an effort to deceive other experts. That Hearnshaw was forced to concoct a speculative psychopathology to make such a fraud charge credible shows the weakness of the charge, which appears even more bizarre as Hearnshaw relied mainly on Burt’s “mixed ancestry” — part Saxon and part Celtic.

It is relevant that several subsequent independent studies gave the same results as those of Burt. (The most recent, the Minnesota Twin Study, was published in *Science*, 12 October 1991, too late for inclusion in Fletcher’s book.) Also important is a 1972 statement by 50 eminent scientists, including several Nobel laureates, praising Burt’s posthumous Thorndike award article for drawing attention to the great influence of heredity in human behaviour. The statement praised Burt’s courage because scientists were suffering suppression, censure, punishment and defamation for emphasizing the role of heredity, although such influences are well documented. Unfortunately, this situation continues.

A fair appraisal of all the circumstantial evidence presented by Fletcher compels the conclusion that Burt was eccentric and sometimes careless, but was guilty of heresy rather than fraud. Fletcher is also warranted in asserting that the issues in the Burt case go beyond the vindication of an individual and show the danger of the intrusion of politics into science in skewing or suppressing research results. Fletcher’s legalistic style is not ingratiating, but for those concerned with psychology or with the integrity of science this is an important and persuasive book. □

Lee Loevinger is at Hogan & Hartson, 555 Thirteenth Street NW, Washington DC 20004, USA.

New In paperback

■ Harvard University Press have recently published *Night Thoughts of a Classical Physicist* by Russell McCormmach. This fascinating experiment in historiography is a portrait of a composite physicist, a man who never existed but whose career, outlook and activities are based upon those who did. Price is £7.95.

■ *Cell Communication in Health and Disease* edited by Howard Rasmussen is an addition to the series ‘Readings from Scientific American’. The collection of essays surveys our understanding of cellular signalling, second messengers and signal transduction. Publisher is Freeman, price is £10.95, \$13.95.

Predominance of long-wavelength heterogeneity in the mantle

Wei-jia Su & Adam M. Dziewonski

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

Analysis of travel-time anomalies of long-period shear waves recorded by the Global Digital Seismographic Network indicates that mantle heterogeneities, which probably reflect patterns of mantle convection, occur most prominently on length scales greater than about 6,000 km, when projected onto a two-dimensional map at the Earth's surface. Models of mantle convection will need to be able to produce a similar power spectrum.

SEISMIC wave speeds change with temperature and composition; so does the Earth's density. It was expected, therefore, from the beginning of global seismic tomography (imaging in three dimensions of small deviations of wave speeds from a spherically symmetric reference model) that the method could provide the critical evidence needed for an improved understanding of the Earth's global dynamics. In 1977, Dziewonski *et al.*¹ wrote: "Lateral variation of the physical properties of the matter within the earth must be the cause as well as the result of the dynamic processes occurring in its interior. There is, as yet, no consistent explanation of the driving mechanism of the plate motions . . . Knowledge of the three-dimensional perturbations of the physical parameters could provide clues to the nature of the mechanism as well as constraints for its quantitative solution." Recently, Olson *et al.*² have described the remarkable convergence of results in seismic tomography, mantle geochemistry and numerical modelling of mantle convection. They present a picture of whole-mantle convection dominated by large-scale flow which is consistent with most of the currently available results.

But the distribution of power in the spectrum of mantle heterogeneity obtained by seismic methods has not yet been satisfactorily investigated. In some models, the spherical harmonic expansion is truncated *a priori* at a low degree; for example, degree two³, four⁴, six⁵⁻⁷ or eight⁸. Alternatively there is a class of models represented by tens of thousands of unknown parameters^{9,10}. The nominal horizontal resolution of these models reaches degrees as high as $l=36$ in the spherical harmonic expansion, and they can have a nearly white spectrum¹¹. There are also uncertainties in modelling of mantle convection; to make computation feasible, it is carried out for Rayleigh numbers two orders of magnitude lower than thought to be appropriate². Consequently, one could expect features at higher wavenumbers in the actual convection pattern than those predicted by calculations (for example in ref. 12). To make proper use of the information provided by seismic tomography and modelling of convection, we must know the spectral distribution of lateral heterogeneity in the mantle.

Worse still, the low-order models may be biased, as well as being incomplete, because of the uneven distribution of sources and receivers, if there is significant power in the rejected part of the spectrum. Two recent papers suggest that this may be the case. Gudmundsson *et al.*¹³ presented results of a statistical analysis of the P-wave travel-time residuals and concluded that in the upper mantle the spectrum of lateral heterogeneity is

white, down to wavelengths as short as 100–200 km. Snieder *et al.*¹⁴ suggested that what seem to be large-scale features in both upper- and lower-mantle tomographic models may be an artefact of low-pass filtration of relatively narrow structures such as mid-ocean ridges or subduction zones. In reality, they say, the highly anomalous properties are limited to narrow zones, whose width is an order of magnitude less than indicated by the tomographic models. If either the above statements is correct, then much of the tomographic work may be wrong or, at best, misleading. In turn, this could invalidate conclusions such as those of Olson *et al.*², which convey the impression that the basic question of geodynamics (the driving mechanism of plate tectonics) has been conceptually answered.

Seismic tomography involves a quantitative and qualitative leap from consideration of one-dimensional Earth structure. For example, Dziewonski and Anderson¹⁵ used four unknowns to

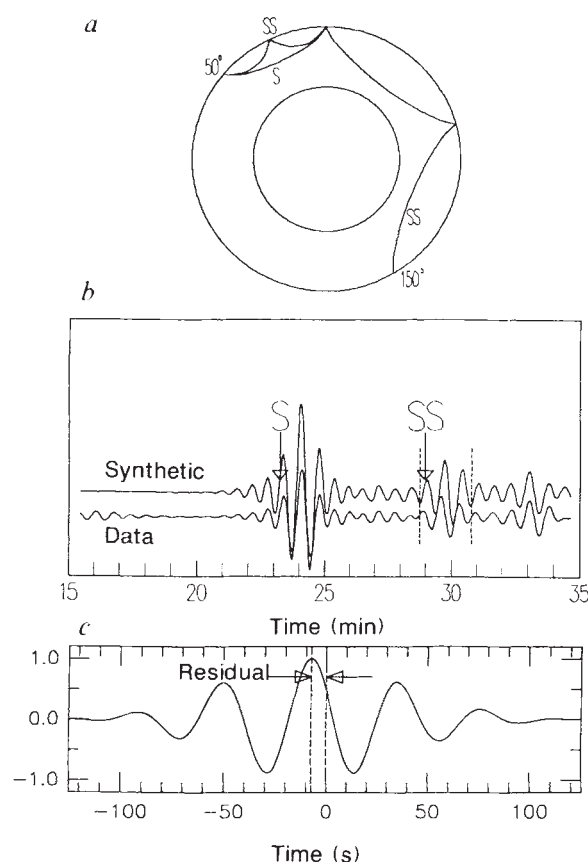


FIG. 1 *a*, S and SS ray paths for the distance 50° (left), and SS ray path for the distance 150° (right). *b*, A comparison of an observed seismogram (24 March 1984) with the corresponding synthetic seismogram calculated for the spherical Earth model PREM¹⁵. Within the dashed lines is the isolated SS phase. *c*, The correlogram for the SS phase in the observed and synthetic seismograms. The offset between the maximum of the correlogram function and time $t=0$ indicates a -7.4 -s residual between the observed and predicted travel times, meaning that the observed SS arrives earlier than predicted by PREM.

TABLE 1 Coefficients of spherical harmonics

		Upper mantle										Lower mantle									
		SS residuals		k=0		k=1		k=2		k=3		k=0		k=1		k=2		k=3		k=4	
l	m	A	B																		
0	0	-1.15		-0.29		1.30		-0.23		-0.10		-0.64		-0.57		0.41		0.02		1.72	
1	0	-1.70		3.30		2.60		0.53		0.93		0.73		-0.44		-0.26		-0.13		-0.14	
1	1	-1.00	-1.23	4.78	2.03	1.42	1.30	-0.23	1.30	-0.21	0.53	-0.50	0.18	0.17	-0.15	0.03	0.39	-0.26	0.26	-0.12	0.15
2	0	-0.68		1.02		2.28		-0.48		-0.82		0.64		-1.62		0.02		0.49		-0.42	
2	1	-0.16	0.11	-0.84	0.32	0.77	-1.56	0.66	-2.47	-0.66	-2.37	-0.32	0.30	0.26	-0.27	-0.04	-0.30	0.24	-0.26	0.05	-0.31
2	2	0.12	1.19	1.61	-4.06	1.04	0.63	1.72	0.64	0.83	0.52	-2.03	-1.16	0.99	-0.44	0.05	-0.34	0.17	0.30	0.06	-0.18
3	0	0.17		0.43		0.12		-0.20		0.13		-0.86		0.37		0.24		0.24		-0.05	
3	1	-0.04	-0.14	1.07	0.73	0.44	0.22	0.70	-0.35	0.20	-0.13	-1.03	-0.42	0.46	0.21	0.03	-0.27	0.27	-0.09	0.15	0.00
3	2	0.45	-0.71	-0.67	1.73	-1.44	2.02	-0.17	0.50	-0.34	0.24	-0.23	1.07	0.09	0.30	0.25	-0.02	0.30	-0.50	0.29	-0.32
3	3	-0.17	-0.66	0.50	1.26	-1.66	-0.18	-1.55	-0.96	-0.44	0.10	-0.69	0.07	-0.17	-0.06	0.09	0.20	0.00	0.25	0.18	-0.19
4	0	-0.33		1.36		1.70		0.80		-0.15		-0.22		0.26		-0.21		-0.17		-0.05	
4	1	0.13	0.29	-0.45	0.03	0.52	0.72	1.19	0.55	0.24	0.05	-0.14	-0.14	-0.18	-0.67	-0.06	0.05	0.08	0.32	-0.16	0.33
4	2	0.52	-0.49	-2.09	0.68	-1.24	1.15	0.23	0.00	0.66	-1.13	0.38	-0.29	-0.24	-0.06	0.04	-0.31	-0.17	-0.24	-0.10	0.05
4	3	-0.03	-0.18	0.03	0.04	-1.28	0.24	-0.49	-0.01	-0.06	0.60	-0.23	0.62	0.30	0.22	0.39	0.00	-0.01	0.14	0.13	0.16
4	4	-0.28	-0.55	-0.06	3.41	0.98	1.95	0.39	1.39	0.60	-0.25	0.24	0.59	0.38	-0.21	0.18	0.12	-0.15	-0.42	-0.50	0.03
5	0	0.50		-2.41		-1.87		0.11		0.33		0.35		-0.18		0.19		0.08		0.22	
5	1	-0.22	-0.28	2.03	0.19	1.32	0.16	0.00	0.65	-0.80	-0.35	-0.01	-0.38	-0.20	0.15	0.17	0.59	0.22	0.48	0.15	0.27
5	2	0.04	0.15	-1.97	0.44	-2.44	-0.45	-0.68	-0.66	0.18	-0.13	0.42	-0.46	0.08	-0.25	0.21	0.19	0.12	0.48	-0.12	-0.10
5	3	-0.23	-0.05	-0.82	1.55	-0.20	1.76	-0.22	-0.15	0.42	-0.22	-0.18	0.12	0.10	0.28	-0.28	0.22	-0.33	0.19	-0.25	0.20
5	4	-0.46	0.40	3.06	-2.38	2.89	-2.81	0.48	-0.76	-0.37	0.03	0.09	0.40	-0.27	0.45	0.02	0.23	0.13	-0.17	0.13	0.17
5	5	-0.69	0.18	1.11	0.90	0.55	0.54	0.19	0.01	0.63	-0.35	-0.22	0.04	0.17	0.20	0.10	0.11	0.19	0.25	-0.13	0.19
6	0	-0.42		0.40		0.91		0.58		0.48		-0.09		-0.17		-0.16		-0.19		0.17	
6	1	0.20	0.19	-0.61	-0.91	-0.35	-0.57	0.08	-0.14	-0.26	0.24	-0.04	-0.24	-0.04	-0.02	0.19	0.03	0.09	0.00	-0.02	0.09
6	2	0.59	0.10	-1.08	0.39	-1.00	1.34	-1.02	1.36	0.19	0.09	-0.13	0.01	-0.05	-0.52	-0.10	-0.10	-0.01	0.08	0.14	0.00
6	3	0.17	-0.69	-0.08	2.22	0.24	0.78	-0.22	-0.37	0.53	-0.26	0.21	-0.10	0.12	0.36	0.01	-0.40	0.07	-0.19	0.36	0.43
6	4	-0.16	-0.05	0.22	0.98	0.91	0.05	0.99	0.02	0.50	0.35	0.57	-0.16	0.03	-0.45	-0.06	-0.10	-0.25	0.15	-0.06	0.32
6	5	-0.50	0.31	1.15	0.50	1.28	0.59	0.46	0.16	-0.55	-0.56	-0.49	-0.34	0.19	-0.19	-0.13	-0.30	0.10	-0.17	0.34	0.16
6	6	-0.34	-0.02	0.25	0.88	-0.54	0.61	-1.01	-0.03	-0.44	0.11	0.22	-0.09	0.17	0.10	0.24	-0.12	-0.21	-0.09	-0.24	0.03
7	0	0.14		-1.00		-0.64		0.15		0.31		-0.40		0.17		0.08		0.08		0.17	
7	1	-0.05	-0.30	-0.30	1.54	-0.22	1.12	-0.16	0.34	-0.42	-0.38	-0.55	0.34	-0.08	0.05	-0.05	-0.08	0.04	-0.24	0.13	0.15
7	2	0.12	-0.14	-0.10	1.52	0.41	1.30	0.40	0.77	-0.03	0.26	-0.06	-0.18	0.65	-0.09	0.22	-0.02	0.19	0.17	0.36	0.24
7	3	0.12	0.13	-0.22	1.34	-1.06	0.68	-0.98	-0.71	0.13	0.40	0.25	-0.62	-0.04	-0.38	0.22	0.13	-0.23	0.41	-0.55	0.11
7	4	0.33	-0.04	0.31	-0.63	-0.69	0.78	-0.83	0.17	0.25	0.36	-0.24	0.28	-0.08	-0.12	-0.06	-0.09	0.25	-0.34	0.48	-0.37
7	5	0.18	0.35	0.55	-0.72	0.23	-0.45	0.09	-0.85	-0.32	0.21	-0.42	-0.16	-0.11	-0.29	-0.02	-0.16	-0.15	-0.34	0.01	-0.06
7	6	0.00	0.12	0.08	-1.15	0.06	-1.40	-0.27	-0.30	-0.56	-0.07	-0.70	0.31	-0.08	-0.26	-0.21	0.05	-0.24	-0.12	0.17	-0.59
7	7	-0.21	-0.06	-0.93	0.06	-0.68	-0.55	-0.35	0.28	0.38	0.11	-0.37	-0.19	-0.10	-0.10	-0.14	0.14	0.17	0.40	0.14	-0.06
8	0	0.50		-1.04		-1.11		-0.37		-0.12		0.00		0.37		-0.02		-0.07		-0.11	
8	1	0.06	0.17	-0.72	-0.40	-0.26	-0.02	-0.32	-0.11	-0.52	-0.32	0.25	0.31	0.19	0.36	0.07	-0.04	0.06	-0.02	0.04	-0.05
8	2	-0.19	-0.18	-0.78	0.35	-1.01	0.00	-1.04	0.05	-0.50	0.00	-0.12	0.21	0.12	-0.16	-0.18	0.13	-0.32	0.08	0.01	0.11
8	3	0.36	-0.32	0.25	0.36	0.84	0.06	0.93	-0.43	0.88	0.02	-0.24	0.28	0.24	0.42	0.18	0.18	0.23	0.21	-0.17	0.21
8	4	-0.01	0.23	-0.31	1.21	0.58	0.23	0.53	-0.07	0.33	0.02	0.49	-0.09	-0.23	-0.06	-0.23	0.17	-0.38	0.23	-0.08	0.00
8	5	0.16	0.03	-0.73	-0.06	-0.63	0.25	0.19	0.31	-0.23	0.44	0.08	0.00	-0.09	-0.40	0.03	0.14	-0.20	0.23	-0.01	0.08
8	6	0.14	-0.08	-0.38	0.68	-0.28	-0.31	0.55	-0.16	-0.28	-0.45	0.34	0.05	-0.01	0.24	-0.28	0.05	0.11	0.10	0.27	0.06
8	7	-0.26	-0.07	0.85	-0.07	0.07	0.27	-1.31	0.10	-0.85	0.18	-0.20	-0.33	0.08	-0.48	0.08	0.03	0.12	0.37	-0.34	-0.06
8	8	-0.04	0.04	0.14	0.27	-0.39	0.26	0.04	-0.51	-0.24	-0.11	0.09	0.33	0.01	0.04	-0.01	-0.14	-0.08	0.03	0.21	0.22

The first column gives the coefficients of spherical harmonics up to degree $L=8$ of the full ($L=36$) expansion of travel-time residuals which fitted the data with a r.m.s. error 1 s. The travel-time residual for these coefficients is $\delta t = \sum_{l=0}^L \sum_{m=0}^l (A_l^m \cos m\phi + B_l^m \sin m\phi) p_l^m(\theta)$, with p_l^m defined in equation 1. Columns 2–10 are spherical harmonic coefficients of the whole-mantle S velocity model SH425.2. The velocity perturbation is $\delta v_s/v_s = 10^{-3} \sum_{k=0}^K \sum_{l=0}^L \sum_{m=0}^l f_k(x) (A_l^m \cos m\phi + B_l^m \sin m\phi) p_l^m(\theta)$. For definitions of $f_k(x)$, see ref. 5 for the lower mantle and ref. 8 for the upper mantle.

obtain the lower mantle P-velocity in a spherically symmetric PREM model, whereas the low-order model of Dziewonski⁵ used 245 parameters to describe the same region in three dimensions, an increase by a factor of sixty. Even greater increases are typical of the 'block' models: the lower-mantle part of a model of Clayton and Comer⁹ contains a three-dimensional array of 57,000 cells, more than 14,000 times the number of parameters in PREM. But this is a 'mathematical trickery'; the available data do not carry this much independent information. For instance, in their waveform inversion of over 2,000 seismograms, each of several hours duration, Woodhouse and Dziewonski⁸ found that only 192 significant eigenvalues (independent parameters), out of 324 required by their low-order model of the upper mantle, led to a discernible decrease in variance. It seems, therefore, that the data base is insufficient to provide a three-dimensional image with enough lateral and radial resolution to verify the conclusions of Gudmundsson *et al.*¹³ or Snieder *et al.*¹⁴, or to confirm or disprove the validity of the assumptions made in the low-order expansion models. This, we believe, would be true even if the data used so far were supplemented by those recorded by the expanding global network of high-quality seismographic stations and by measurements from projects such as the International Seismic Observing Period¹⁶.

Here we step back to complete the missing stage in the progression of seismology from one to three dimensions by investigating a two-dimensional surface projection of three-dimensional lateral heterogeneity. Seismic tomography may be compared to inferring the three-dimensional shapes of objects from the shadows they cast on a screen (measurements on the Earth's surface). This can be tricky but, with only elementary precautions, a simple inspection of the 'shadows' should allow us to decide whether there are relatively few large objects or many small ones. If the former is true, then the low-order expansion is justified, and a global three-dimensional inversion can be attempted with a good chance of success. If the shadows indicate predominance of small-scale objects or a uniform mixture of sizes (white spectrum), the problem is much more difficult.

An ideal data set for a two-dimensional study might consist of measurements of vertical ScS travel-time anomalies on a uniformly distributed, dense grid of points. In our 'shadow' analogy, the rays are in this case at normal incidence to the screen and the problems resulting from oblique illumination are avoided. One could perform spherical harmonic analysis of such a data set and obtain information on the spectral distribution of lateral heterogeneity without solving a two- or three-dimensional inverse problem. In this way, a much higher

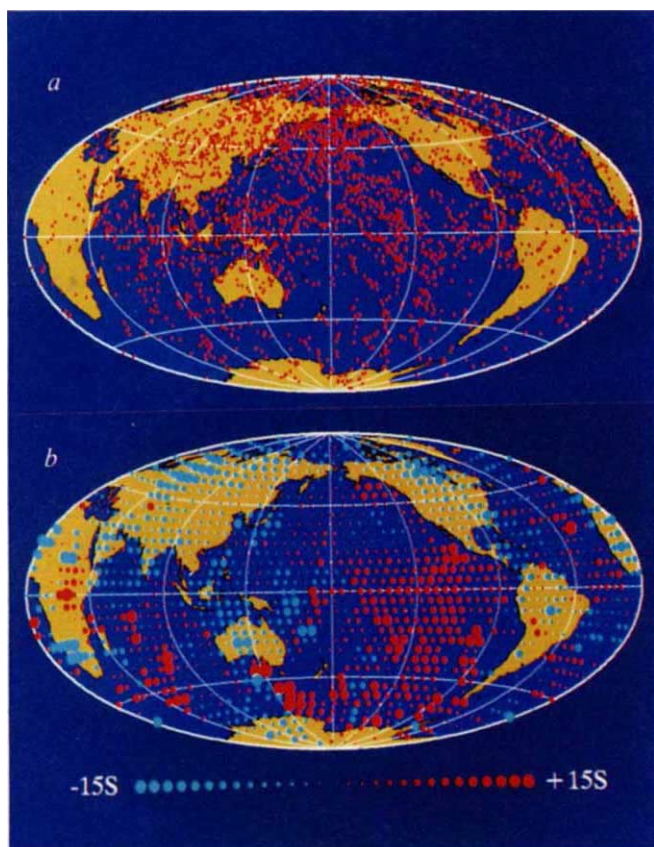


FIG. 2 *a*, Distribution of SS-phase mid-path bounce points at the Earth's surface; each circle represents one of the 3,313 data points in this study. *b*, Travel-time residuals for the SS phase projected on an equally spaced grid ($5^\circ \times 5^\circ$ at the equator). The residuals are smoothed by a moving cap of radius 5° . The values shown are corrected for topography, bathymetry and crystal thickness.

horizontal resolution could be reached even when the smoothing effect of integration along the ray path and the lack of depth resolution are taken into account. This particular experiment is not feasible, because the distribution of earthquakes and stations is far from that needed. But, with minor complications, other seismic phases can be used instead.

The data set

We choose the phase that seismologists call SS. It is a downward-radiated shear wave reflected once at the Earth's surface, roughly

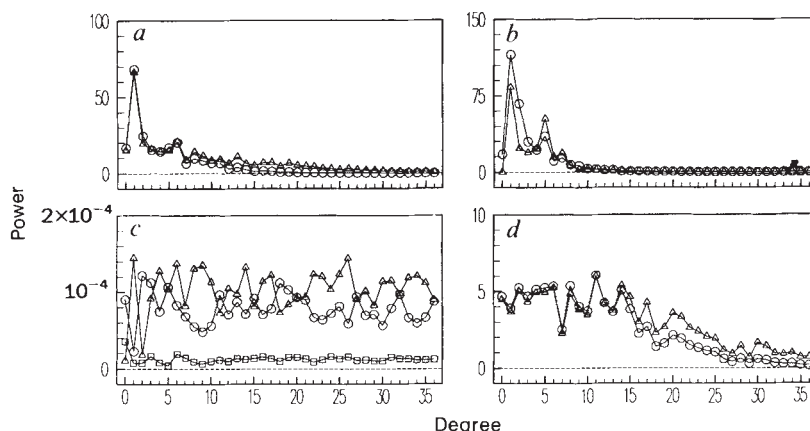
at the mid-point of its path between the source and receiver. We use SS waves recorded in the epicentral distance range from 50° to 150° . The ray paths for these two distances are shown in Fig. 1*a*, in which we also show the S phase at 50° . The depths at which these S-wave rays bottom are between 750 and 2,000 km. This helps to avoid complications caused by the steep gradients and discontinuities in the upper mantle as well as the high level of heterogeneity in the lowermost mantle^{5,17}. Also, we use only the transverse (SH) component of the ground motion. The displacements recorded on the vertical or longitudinal components (P–SV) are much more complex because of the P to S and S to P conversions.

Our data come from the recordings of long-period channels of the Global Digital Seismographic Network. Their response is poor for periods shorter than 15 s and reading of arrival times of individual phases is very inaccurate. Instead, we determine the differences between the recording and the synthetic seismogram computed for the spherically symmetric reference model PREM¹⁵. Figure 1*b* compares an observed seismogram (bottom) with the corresponding synthetic seismogram (top) calculated by superposition of $\sim 11,000$ normal modes¹⁸. Both the observed and synthetic seismograms are low-pass filtered with a cut-off period of 32 s. We obtain the source mechanism required for calculating the synthetic seismogram from the Harvard catalogue, which contains over 8,000 centroid-moment tensor solutions for all events between 1977 and 1989¹⁹ likely to be used for the present purpose. Figure 1*c* shows the cross-correlation between the sections of the observed and synthetic seismograms delineated by the vertical dashed lines. The time at which the cross-correlation function reaches maximum is the measurement sought²⁰.

We selected 429 earthquakes of appropriate size and distribution from the period 1977–89; these yielded 3,313 satisfactory measurements of the SS residuals in the appropriate distance range whose mid-path surface reflections are shown in Fig. 2*a*. The coverage is better in the Northern Hemisphere, but it is adequate for our purpose south of the equator. Figure 2*b* shows the SS residuals projected on a roughly equally spaced grid corresponding to $5^\circ \times 5^\circ$ at the Equator. Following Woodward and Masters²¹, the displayed values are averages of all available data within 5° of a given grid point. The values shown are corrected for the topography, bathymetry and crust thickness²².

There are SS residual values for 1,452 caps, or more than 90% of all the grid points. Even visual examination of Fig. 2*b* shows the anomalies tend to change slowly with position, even though the smoothing introduced by averaging involves only the adjacent grid points. The other conclusion is that there is a clear correlation between the sign of the anomaly and the tectonic nature of the mid-path reflection point. This means that the upper-mantle structure underneath that point is the most obvious contributor to the anomaly. Nevertheless, contributions from the lower mantle are also important²¹; although the r.m.s.

FIG. 3 Power in each harmonic degree, as defined by equation (2), for the spherical harmonic expansions of *a*, different fits to the observed SS travel-time residuals (Δ , exact; $\circ$, damped); *b*, the damped fit to SS travel-time residuals predicted by the upper-mantle model M84C (Δ) and the whole-mantle S-wave model SH425.2 ($\circ$); *c*, the velocity anomalies ($\delta v/v$) at 100 (Δ), 500 ($\circ$) and 1,500 km ($\square$) in the 'random' model, and *d*, SS travel-time residuals predicted by the random structure up to degree 36 (Δ , exact; $\circ$, damped). The power is expressed in units of s^2 , except for (*c*), where the units are dimensionless.



heterogeneity is less than in the upper mantle, the path segments are significantly longer (see Fig. 1a).

Our choice of the SS phase was influenced by the results of Woodward²² and Woodward and Masters^{17,21} who measured travel-time differences between SS and S as well as between ScS and S. The differential travel-time approach has the advantage of being relatively insensitive to the source parameters and to the structure near the source and the receiver. The 'absolute' travel-time approach has the principal advantage of a wider range of epicentral distances in which measurements can be made (50–150°, as compared with 50–100° for SS–S). This is the main reason for our relatively good coverage of the Southern Hemisphere. Also, the interference of other phases such as ScS may be less critical when cross-correlation is made with respect to a synthetic seismogram, as the interfering phase is present in both seismograms.

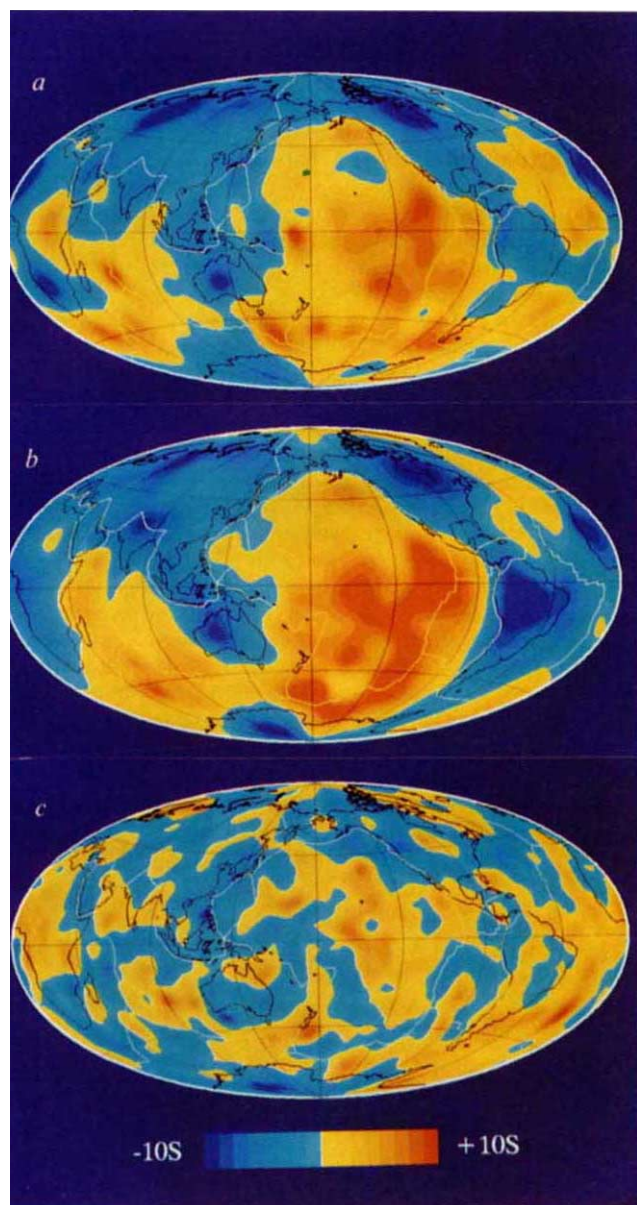


FIG. 4 a, Map of the observed SS travel-time residuals synthesized from spherical harmonic coefficients up to $l=36$. b, As for a, but for residuals predicted by the whole-mantle S-wave model SH425.2. c, As for b, but for the random velocity model. The travel-time residual scale range is ± 10 s.

SS spectra and synthetic tests

We fit spherical harmonics of degrees 0–36 to the data from 1,452 grid points shown in Fig. 2b. The associated Legendre functions are normalized

$$p_l^m = (2 - \delta_{m,0})^{1/2} (2l+1)^{1/2} \left[\frac{(l-m)!}{(l+m)!} \right]^{1/2} P_{lm} \quad (1)$$

so that the r.m.s. value over the surface of the sphere is equal to one for each sine or cosine harmonic. We use spherical splines²³, although with coverage as complete as that shown in Fig. 2b, a direct least-squares fit would be stable. In choosing the accuracy of the fit to the data, we consider two cases. First, the data should be fitted as accurately as possible: this is the 'exact' solution. Second, the data should be fitted to give a r.m.s. error of 1 s, which seems conservative considering the potential sources of error: this is the 'damped' solution. The coefficients for the latter solution are listed in Table 1 for the first eight orders. Figure 3a shows the sums

$$D_l = \sum_{m=0}^l A_{lm}^2 + B_{lm}^2 \quad (2)$$

for both solutions. There is very little power in the 'damped' solution beyond about degree 10. The power decreases less rapidly for the 'exact' fit, but even in this case it approaches zero for degrees greater than 20.

It is instructive to calculate the SS residuals for existing models of the Earth and compare their spectra with those in Fig. 3a. We use two Earth models: the first is the upper-mantle model M84C⁸; the other is a new whole-mantle shear-velocity model SH425.2. It is based on the same data set as model U84L85/SH^{24–26}, but subsets of data have been weighted differently and the model is less damped. The number of eigenvalues used is 425, out of 729, and three iterations of the nonlinear inverse problem were completed. The crustal correction is that used in ref. 8. The radial variations are described, as in model U84L85/SH, by the Legendre polynomials up to degree 3 in the upper mantle and degree 4 in the lower mantle; the spherical harmonic expansion is truncated at degree eight in both regions. The coefficients of model SH425.2 are listed in Table 1. The SS residuals are calculated using the ray-tracing algorithm of Julian and Gubbins²⁷ for exactly the same source-receiver pairs for which we have observations, and the residuals have been averaged at the same grid points as shown in Fig. 2b. The power spectra for both models are shown in Fig. 3b. It is clear that the overall distribution of power in the 'damped' model is very similar to that predicted by the three-dimensional models of the mantle; significant differences exist only for degree 1 for model SH425.2 and degree 5 for M84C.

Integration of anomalies along a ray path is a smoothing process, however, and one could still suppose that the actual structure has a substantial power at high orders. To answer this question, we have created a velocity model in which the anomalies are random at each depth and the r.m.s. perturbation changes with depth in the same way as the power in model SH425.2. The random variations have been scaled such that at each depth the expected total power for each degree should remain constant. This is shown in Fig. 3c for three different depths.

Using this hypothetical model, we calculated the SS residuals in exactly the same way as described above, and Fig. 3d shows the spectral power for the 'damped' and 'exact' spherical harmonic fits. Calculating the SS residuals for the random structure is equivalent to evaluating the response of a complex filter. If the filter were 'all-pass', the spectra in Fig. 3d would be roughly flat for all orders from 0 to 36. Instead, the response is 'all-pass' up to $l \approx 15$, and then it begins to decrease gradually. This decrease reflects the smoothing effect of integration along the ray path and averaging of data over the spherical caps. These

spectra are distinctly different from the observed spectra in Fig. 3a, where the abrupt decrease in power occurs at $l=6$. The synthetic spectra have much more energy at high angular order numbers, so that the smoothing through integration is not sufficient to mask the presence of high-wavenumber structure.

Discussion

Figure 4a is a map derived by the synthesis of the spherical harmonic coefficients for degrees 1–36 obtained in a ‘damped’ fit to the observed SS residuals. Figure 4b is the same as 4a but for the residuals predicted by model SH425.2. Figure 4c is also the same as 4a, but for the ‘random’ model. This figure reflects, roughly, the pattern of the SS residuals appropriate for an upper mantle with a spectrum of lateral heterogeneity such as proposed by Gudmundsson *et al.*¹³. The spectral content of this prediction is very different from that of the observations in Fig. 4a.

The similarity between Fig. 4a and b is striking: the correlation coefficient for degrees 1–36 is 0.74, but it is as high as 0.86 for degrees 1–6. The correlation coefficient is above 0.79 for each degree from 1 to 6; it drops suddenly to 0.29 for degree 7 and to 0.20 for degree 8, indicating that the power beyond degree 6 might be too low to be resolved on the global scale. It should be pointed out that the SS data were not explicitly used in deriving model SH425.2. Even though ~60% of events used were the same, the method of Woodhouse and Dziewonski²⁴ uses entire waveforms, which consist of mantle waves and many other body wave phases in addition to SS (see Figs 3 and 5 in Dziewonski and Woodhouse²⁵). With the principal difference between the observed and predicted residuals limited to a region of the East Pacific Rise and South America, the agreement indicates that the ‘average path approximation’⁸ works rather well not only for surface waves but also for body waves.

The principal conclusion to be drawn from Figs 3 and 4 is that most of the power in the lateral heterogeneity spectrum

with wavelengths greater than ~1,000 km is contained in angular order numbers 6 and below, which correspond to wavelengths greater than 6,000–7,000 km at the surface. Only in this way could a low-order model such as SH425.2 predict so well the observations at much shorter resolving lengths. This means that important features of the early tomographic models such as the degree 2 anomaly in the transition zone³ or the ring of high velocities circumscribing the Pacific basin^{1,5} are not artefacts of low-pass filtration of much narrower structures¹⁴, but a reflection of the actual properties of the Earth.

This conclusion is supported by four other studies, the first two of which mainly bear on the properties of the upper mantle. Woodward and Masters²¹ investigated SS–S residuals, and although their geographical coverage of the mid-path surface reflection points is sparser than ours, particularly in the Southern Hemisphere, the dominating influence of the large-scale heterogeneities is clear both in the raw and in the smoothed data. Zhang and Tanimoto²⁸ conducted a high-resolution study of a global distribution of the Love-wave phase velocities for waves of period 80–200 s. They found that the spectrum of the ‘local’ phase velocities²⁹ is dominated by low-order spherical harmonics. Their conclusions are fully compatible with those presented here, even though the data used were different.

The third study is also by Woodward and Masters¹⁷. Their pattern of ScS–S residuals indicates the paramount influence of large-scale heterogeneity. Because the paths of ScS and S in the upper mantle are similar, the residuals must originate in the lower mantle. Through an ingenious test³⁰, they demonstrate that the principal contribution comes from depths greater than 2,000 km, as predicted by model L02.56 (ref. 5). The fourth study is different in nature, because it involves a very-large-scale inversion. Inoue *et al.*¹⁰ use the P-wave travel-time anomalies taken from the bulletins of the International Seismological Centre to obtain velocity perturbations in 16 spherical shells spanning the whole mantle, each divided into 32×64 , or 2,048, blocks ($5.6^\circ \times 5.6^\circ$) for a total of 32,768 unknown parameters. The problem obviously has too many parameters but a stable, converging solution is obtained by using smoothing criteria. The authors kindly made their model available to us, and we have found the spherical harmonic coefficients up to degree 36 for each shell. Figure 5 shows power spectra for two shells: one in the upper mantle, the other in the lower mantle. In both cases there is very little power for degrees greater than 6; thus 2,048 parameters in each of the two shells are well represented by only 48 spherical harmonic coefficients. We found similar results for the other shells.

Of course, smaller-scale features do exist. This is certainly true near the Earth’s surface, where for example, sharp boundaries between the continental and oceanic structures cannot be effectively described by the low-order expansion. Correlations between the distribution of hot spots and large-scale low-velocity anomalies near the core–mantle boundary (see plate 2g in ref. 26) indicate that there might be a relationship, perhaps non-linear, between the large- and small-scale features.

The expansion of the plate velocity field in terms of spherical harmonic coefficients, with the spheroidal and toroidal components, may serve as a useful analogy, as the random measurement errors are not involved. Hager and O’Connell³¹ found that the power spectrum of these coefficients decreases rapidly, roughly as l^{-2} , not unlike the SS spectrum: The explanation is that the velocity field is dominated by very few large plates with nearly constant velocities, thus giving rise to low-degree energy. The discontinuities are represented by a spectrum that decreases as l^{-2} in the two-dimensional case. Although the individual coefficients at high l are small, their effect can be large on summation because they are in phase.

We think that our result can be interpreted as follows. The field of velocity anomalies is dominated by entities whose horizontal dimensions are several thousand kilometres. Velocity anomalies change smoothly within these entities. The velocities

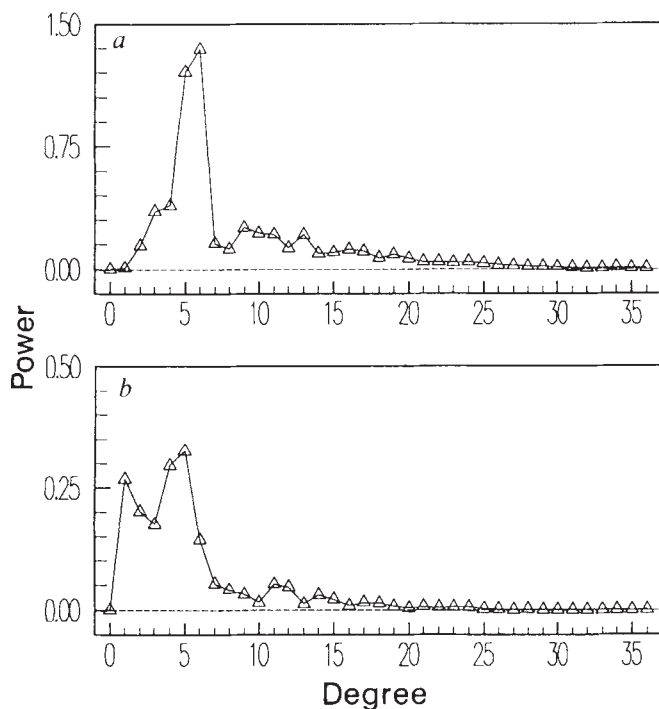


FIG. 5 Power in each harmonic degree for the spherical harmonic expansions of $\delta v_p/v_p$ given by the whole-mantle P-velocity model of Inoue *et al.*¹⁰. a, Layer 4 (148–238 km), b, layer 16 (2,566–2,900 km).

may change abruptly at the boundaries. Also, there may be a finite number of smaller-scale anomalies that are planar in shape (subduction zones) or linear (plumes), but the total power associated with these features is insufficient to bias the results of a properly carried out 'low-order' inversion.

These results impose constraints on modelling of mantle convection. The experiments of Hager *et al.*³² and Forte and Peltier³³ demonstrate that seismic velocity anomalies are principally caused by differences in temperature. (The presence of chemical heterogeneity cannot be ruled out, but its effect on lateral variations in velocity seems to be minor throughout most of the mantle, with the possible exception of the D'' layer.) Assuming that perturbations in velocities are proportional to those in temperature, the temperature field of a successful mantle convection model integrated along the SS ray paths, for example, should be characterized by a spectrum similar to that in Fig. 3a.

Some final remarks with regard to seismological practice. The main reason why our experiment and those of Woodward and

Masters^{17,21} or Zhang and Tanimoto²⁸ have succeeded is that they were performed at relatively long periods. With a cut-off period of 32 s, each of our SS ray paths samples an average volume at least 1,000 times as large as 1 Hz frequency, which is characteristic of the P-waves reported in the bulletins of the International Seismological Centre. This automatically eliminates most of the noise associated with small-scale structures. Explanation of a spectrum from a single source-station pair may, for a low-order normal mode, require integration over the entire volume of the Earth³⁴. Long-period, large-wavelength data are very effective volume integrators, and their availability is essential for studying the Earth on a global scale. Our experiment defines this scale to be confined to degrees less than 10. A global network of about 100 well distributed seismographic stations equipped to provide high-quality long- and very-long-period data should be sufficient for this purpose. It is encouraging that the deployment of such a network is under way, and that there is even some prospect of permanent stations emplaced in boreholes in the deep ocean bottom³⁵. □

Received 2 April; accepted 18 June 1991.

1. Dziewonski, A. M., Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **82**, 239–255 (1977).
2. Olson, P., Silver, P. G. & Carlson, R. W. *Nature* **344**, 209–215 (1990).
3. Masters, G., Jordan, T. H., Silver, P. G. & Gilbert, F. *Nature* **298**, 609–613 (1982).
4. Morelli, A. & Dziewonski, A. M. *Nature* **325**, 678–683 (1987).
5. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
6. Nataf, H.-C., Nakanishi, I. & Anderson, D. L. *J. geophys. Res.* **91**, 7261–7307 (1986).
7. Tanimoto, T. *Geophys. J. Int.* **100**, 327–336 (1990).
8. Woodhouse, J. H. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
9. Clayton, R. W. & Comer, R. P. *EOS Trans. Am. geophys. Un. (Abstr.)* **64**, 776 (1983).
10. Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y. *Phys. Earth planet. Inter.* **59**, 294–328 (1990).
11. Clayton, R. W. & Comer, R. P. *Terra Cognita* **4**, 282 (Abstr.) (1984).
12. Bercowski, D., Schubert, G. & Glatzmaier, G. A. *Science* **244**, 950–955 (1989).
13. Gudmundsson, O., Davies, J. H. & Clayton, R. W. *Geophys. J. Int.* **102**, 25–43 (1990).
14. Snieder, R., Beckers, J. & Neele, F. *J. geophys. Res.* **96**, 501–515 (1991).
15. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
16. Engdahl, E. R. *EOS Trans. Am. geophys. Un.* **70**, 1501 (1989).
17. Woodward, R. L. & Masters, G. *Nature* (in the press).
18. Dziewonski, A. M., Chou, T.-A. & Woodhouse, J. H. *J. geophys. Res.* **86**, 2825–2852 (1981).
19. Dziewonski, A. M., Ekström, G., Woodhouse, J. H. & Zwart, G. *Phys. Earth planet. Inter.* **62**, 194–207 (1990).
20. Su, W.-J., Dziewonski, A. M. & Woodhouse, J. H. *EOS Trans. Am. geophys. Un. (Abstr.)* **71**, 556 (1990).

21. Woodward, R. L. & Masters, G. *J. geophys. Res.* **96**, 6351–6377, (1991).
22. Woodward, R. L. thesis, Univ. of California (1989).
23. Shure, L., Parker, R. L. & Backus, G. E. *Phys. Earth planet. Inter.* **28**, 215–229 (1982).
24. Woodhouse, J. H. & Dziewonski, A. M. *EOS Trans. Am. geophys. Un.* **67**, 307 (1986).
25. Dziewonski, A. M. & Woodhouse, J. H. *Science* **236**, 37–48 (1987).
26. Woodhouse, J. H. & Dziewonski, A. M. *Phil. Trans. R. Soc. Lond. A328*, 291–308 (1989).
27. Julian, B. R. & Gubbins, D. *Geophys. J. R. astr. Soc.* **43**, 95–113 (1977).
28. Zhang, Y.-S. & Tanimoto, T. *Phys. Earth planet. Inter.* **66**, 160–202 (1991).
29. Jordan, T. H. & Lynn, W. S. *J. geophys. Res.* **79**, 2679–2685 (1974).
30. Jordan, T. H. *Geophys. J. R. astr. Soc.* **52**, 441–455 (1978).
31. Hager, B. H. & O'Connell, R. J. *Tectonophysics* **50**, 111–133 (1978).
32. Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. *Nature* **313**, 541–545 (1985).
33. Forte, A. M. & Peltier, W. R. *J. geophys. Res.* **92**, 3645–3679 (1987).
34. Giardini, D., Li, X. & Woodhouse, J. H. *Nature* **325**, 405–411 (1987).
35. Purdy, G. M., Dziewonski, A. M., Orcutt, J. A., Kanamori, H. & Duennebie, F. K. *EOS Trans. Am. geophys. Un. (Abstr.)* **70**, 1230 (1989).

ACKNOWLEDGEMENTS. Much of our data set and the analytical tools were developed in collaboration with J. Woodhouse. We thank R. Woodward for his computer programs and critical reading of the manuscript, and H. Inoue for numerical details of his model. We thank the staff of the Albuquerque Seismological Laboratory of the US Geological Survey for providing data from the Global Digital Seismographic Network. This work was supported by the NSF.

Actin microfilament dynamics in locomoting cells

Julie A. Theriot & Timothy J. Mitchison

Departments of Biochemistry and Biophysics and Pharmacology, University of California, San Francisco, San Francisco, California 94143-0448, USA

The dynamic behaviour of actin filaments has been directly observed in living, motile cells using fluorescence photoactivation. In goldfish epithelial keratocytes, the actin microfilaments in the lamellipodium remain approximately fixed relative to the substrate as the cell moves over them, regardless of cell speed. The rate of turnover of actin subunits in the lamellipodium is remarkably rapid. Cell movement is directly and tightly coupled to the formation of new actin filaments at the leading edge.

THE motility of individual cells is critically important to many basic biological processes, such as embryonic development, wound healing and inflammation. Many different types of cells, including amoebae, leukocytes, fibroblasts, epithelial cells and neurite growth cones, move by crawling across solid substrates

in a similar process. Although polymerized actin filaments are necessary for this type of locomotion, the mechanisms of force transduction and movement are still unknown.

Much progress has been made in the study of actin biochemistry *in vitro*¹, but it has been difficult to relate this data to the function of actin *in vivo*. Some structural information is available on actin organization^{2,3}, but relatively little is known about the dynamic behaviour of actin in moving cells. Information on *in vivo* actin dynamics will be crucial for linking the available biochemical data to morphogenesis and function.

Polymerized actin seems to move continuously centripetally in motile cells. Monomeric actin apparently polymerizes preferentially at or near the membrane at the extreme leading edge of the cell and the cytoskeleton is then transported backwards towards the cell nucleus^{4–7}. Observations of actin filament movements *in vivo* have usually been performed on cells which were immobile during observation. Thus the relationships among rearward actin flux, protrusion of the leading edge, and translocation of the bulk cytoplasm and nucleus are unclear.

We have addressed this issue by examining the dynamic behaviour of polymerized actin microfilaments in the lamellipodium of locomoting cells using the recently developed technique of photoactivation of fluorescence. This technique has been used in our laboratory to examine the movements and dynamics of microtubules in the mitotic spindle *in vivo*⁸ and *in vitro*⁹. Traditional fluorescent derivatives of cytoskeletal proteins *in vivo* can reveal the steady-state distribution of these proteins, but through photoactivation of a spatially defined subset of cytoskeletal elements, we can observe movements of these elements within the steady-state structure *in vivo*, and quantitate dynamic parameters.

Epithelial keratocytes of goldfish (*Carassius*) were chosen as a system for *in vivo* observations of actin movements. Keratocytes move rapidly using a mechanism that is dependent on actin filaments but independent of microtubules, and they have a uniquely large, homogeneous lamellipodium¹⁰⁻¹³.

We have synthesized a novel 'caged' photoactivatable fluorescent derivative of actin which can copolymerize with underivatized actin *in vitro* and *in vivo*. We have used this derivative to mark actin filaments in a selected region of the lamellipodium of locomoting keratocytes. By observing the fluorescent signal arising from the marked filaments, we can determine the rate of rearward actin filament transport as compared with the rate of forward cell movement. We can also determine the rate of turnover of actin subunits in the lamellipodium. The data increase our understanding of the dynamic organization of the actin cytoskeleton and serve to constrain models for the mechanism of cell locomotion.

Caged resorufin-actin polymerization

Resorufin was identified as a likely fluorochrome for caging because its *O*-alkyl and *O*-acyl derivatives have found wide use

as fluorogenic enzyme substrates¹⁴. Caged resorufin iodoacetamide (CRIA) (Fig. 1a) was synthesized from resorufin iodoacetamide by a modification of the procedure for caging nucleotides¹⁵. Although CRIA absorbs light at visible wavelengths, we found it is at least 100-fold less fluorescent than the parent molecule when excited at 575 nm. On exposure to 360 nm light, CRIA is rapidly and completely converted to resorufin-IA (Fig. 1b).

CRIA was used to covalently label actin to a stoichiometry of 0.65 molecules CR (caged resorufin) to 1 molecule actin. Although we have not determined the site of modification, it is likely to be cysteine 374 at the actin C terminus, as this residue is the most reactive to a variety of thiol-reactive agents¹⁶. Modification of this residue generally does not affect the ability of labelled actin to polymerize¹⁷, and we purified polymerization-competent labelled molecules by a round of polymerization and depolymerization after labelling.

To test whether CR-actin retained function, we assayed whether it could copolymerize *in vitro* with unlabelled actin. CR-actin was mixed with unlabelled actin at a ratio of 1:4. When this mixture was put under polymerizing conditions by raising the salt concentration and adding magnesium, 81 % of the CR-actin polymerized, compared with 94 % of the unlabelled actin. Thus the derivative retains its ability to copolymerize with native actin. As a control to prove that the modified actin was not simply sticking to polymerized filaments, we deliberately inhibited the ability of CR-actin to polymerize by treating it with diethylpyrocarbonate¹⁸. When carbethoxylated-CR-actin was mixed with unlabelled actin and put under polymerizing conditions, only 21% of the labelled actin was found in polymer.

To test whether CR-actin would also copolymerize with unlabelled actin *in vivo*, we injected CR-actin into PtK2 epithelial cells. The CR-actin was allowed to incorporate for 1 h.

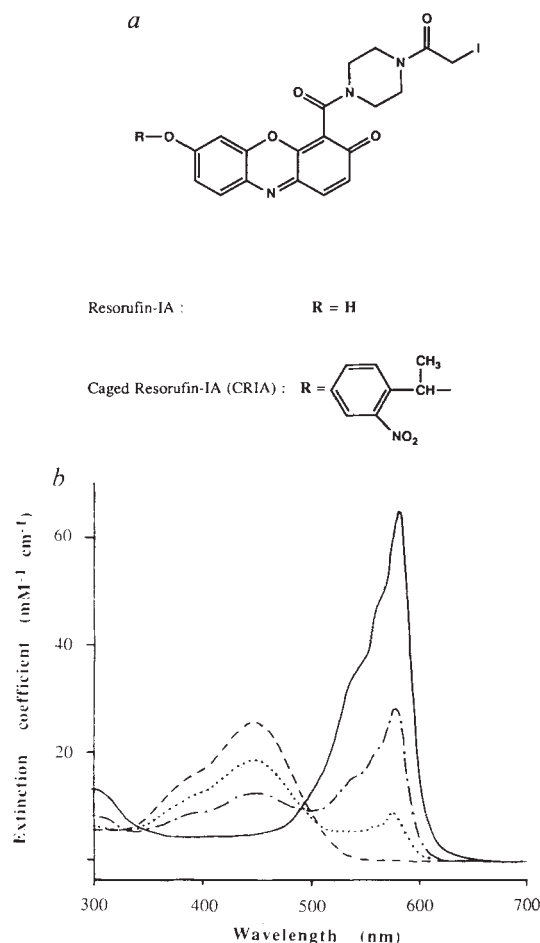


FIG. 1 Structure and activation of CRIA. a, Structure of *N*-(resorufin-4-carbonyl)-*N'*-iodoacetyl piperazine and its caged derivative. b, Optical absorption spectra of caged resorufin after 0 s (—), 30 s (· · · ·), 2 min (- · -) and 10 min (---) of activation in a cuvette with a 360-nm hand-held lamp.

METHODS. All steps were performed under dim orange photographic safe light. 1-(2-Nitrophenyl) diazoethane was prepared according to Walker *et al.*¹⁵. Resorufin-IA (*N*-resorufin-4-carbonyl)-*N'*-iodoacetyl piperazine; Boehringer) (20 μ mol in 100 μ l dimethylformamide; Aldrich Sure-Seal), was added to 80 μ mol of the diazoethane. The mixture was incubated in the dark at room temperature with constant mixing. The progress of the reaction was followed by thin-layer chromatography (TLC) on silica gel (Whatman PE SIL G/UV) in a solvent composed of 67 % benzene and 33 % acetone. Products were visible as two orange spots of retardation factor (RF) 0.48 and 0.58. When the reaction was complete (typically 12–16 h), the products were precipitated with 10 vols ice-cold water and resuspended in a minimal volume of methanol. The products were purified by preparative TLC (Whatman PLK5F) in 67 % benzene, 33 % acetone. The two products were scraped off the TLC plate, eluted with acetone, vacuum evaporated to dryness, and resuspended in dimethylsulphoxide (Aldrich Sure-Seal). The faster migrating product (CRIA-1, RF=0.58) had peak absorption at 450 nm, and the slower (CRIA-2, RF=0.48) at 460 nm. Exposure of both products to 360 nm ultraviolet light from a hand-held lamp resulted in complete conversion to resorufin-IA within 10 min, as judged by analysis of absorption and fluorescence spectrometry before and after photoactivation. We have not determined the structural differences between CRIA-1 and -2, though caging on the two different phenolic oxygens is a possibility. Both products were at least 100-fold less fluorescent than resorufin-IA when excited at 575 nm.

The cells were then fixed and labelled with fluorescent phalloidin to mark stress fibres, and the CR-actin was activated. In injected cells, stress fibres identified by phalloidin staining were coextensive with similar structures labelled with resorufin (Fig. 2a, b). Microinjected CR-actin is therefore competent to assemble into stress fibres *in vivo*.

To investigate the dynamic behaviour of CR-actin in stress fibres, the rate of polymer turnover was measured by activating a small region of a living injected cell and measuring total fluorescence in the stress fibres over time (Fig. 2c-f). The

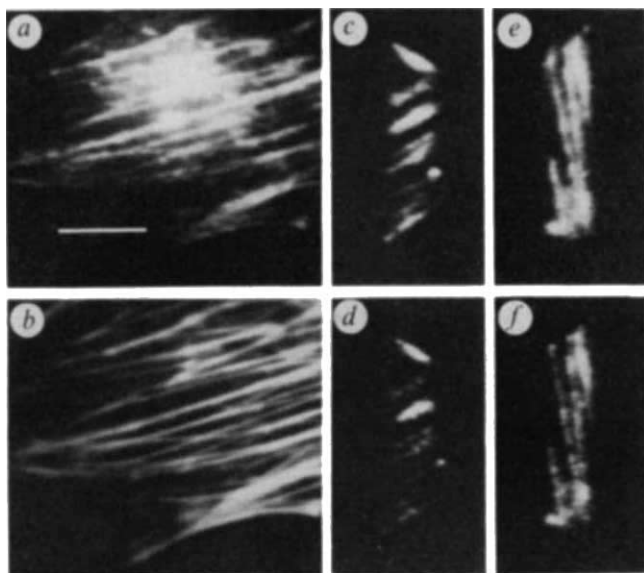


FIG. 2 Incorporation and turnover of CR-actin in PtK2 stress fibres. **a**, CR-actin distribution in a PtK2 potoroo kidney epithelial cell as determined by photo-activation 1 h after microinjection. **b**, Same cell labelled with bodipy-phalloidin. **c**, **d**, CR-actin in an activated region of a living PtK2 cell 5 s (**c**) and 4 min (**d**) after activation. **e**, **f**, Similar cell 5 s (**e**) and 6 min (**f**) after activation. Scale bar, 10 μ m.

METHODS. Rabbit skeletal muscle actin was prepared according to the method of Pardee and Spudich³⁹. CRIA (0.45 mg in 50 μ l DMSO) was added to 1 ml G-actin 5 mg ml⁻¹ in 'G buffer' (5 mM Tris-HCl, pH 8.1, 0.2 mM CaCl₂, 0.2 mM ATP, 10 μ M dithiothreitol (DTT)). After 16 h at 4 °C in the dark with constant mixing, unreacted CRIA was removed by gel filtration over Bio-Gel P-10 polyacrylamide gel (Bio-Rad) in G buffer containing 1 mM DTT. The labelled CR-actin was polymerized by adjusting the buffer to 'F buffer' (50 mM Tris-HCl, pH 8.1, 50 mM KCl, 1 mM ATP, 2 mM MgCl₂, 1 mM DTT). The polymerized actin filaments were collected by centrifugation, resuspended in G buffer with 1 mM DTT, and allowed to depolymerize. Aggregates were removed by centrifugation in an Eppendorf microfuge at maximum speed for 30 min. Protein concentration was determined by optical absorbance. CR concentration was determined by optical absorbance following complete photolysis in a cuvette. The labelling stoichiometry was about 0.65 molecules of CR per actin monomer. PtK2 cells were grown and injected as previously described⁸. Actin was injected at a concentration of 4.0 mg ml⁻¹ in injection buffer (1 mM HEPES, pH 7.5, 0.2 mM MgCl₂, 0.2 mM ATP) and allowed to incorporate in injected cells for 1 h before photo-activation. The volume injected was estimated to be less than 5% of the cell volume. Glutaraldehyde fixation and phalloidin staining were essentially according to the method of Rinnerthaler *et al.*⁴⁰. The optical apparatus and photoactivation beam were as previously described⁸. For imaging of resorufin, we used a 570–590 nm bandpass filter for excitation and a 615-nm long-pass filter for emission. The light source was a mercury arc attenuated with a 12% neutral density transmission filter. The peak of resorufin absorption coincides well with the 577 nm Hg line. The response of the ISIT (Intensified Silicon Intensified Tube) camera was about linear for the light levels typically used in these experiments. Total fluorescence intensity in the stress fibres was measured using Image 1 software, version 3.65 (IVS). Intensity was measured at 30 s intervals and the half-time of turnover was determined by fitting an exponential decay curve to the plot of intensity versus time.

half-life of CR-actin in stress fibres was 3.8 min (s.d. = 1, n = 14). The half-time of turnover of individual stress fibres in a given cell varied by as much as twofold. Thinner stress fibres turned over more rapidly than thicker ones. The average turnover rate is consistent with the rate in fibroblasts measured by incorporation of fluorescently labelled actin after injection (3.3–4.0 min⁻¹), but is faster than the rate measured by fluorescence recovery after photobleaching (about 10 min⁻¹). From our data on its polymerization *in vitro* and incorporation into stress fibres *in vivo*, we conclude that CR-actin is a reasonable probe for following the movement and turnover of actin filaments.

Actin dynamics in lamellipodia

The lamellipodium of goldfish epithelial keratocytes is typically about 5–15 μ m wide and contains a fairly uniform distribution of actin microfilaments as visualized by rhodamine-phalloidin staining. Fluorescence intensity profiles of fixed labelled cells indicate that the actin filament density in the front and rear of the lamellipodium rarely differs by more than about 20% (Fig. 3).

For studies of actin dynamics, keratocytes were cultured on glass coverslips and microinjected with CR-actin. The injected actin was allowed to incorporate for 20 min to 1 h. Only cells that appeared unaffected by the injection and continued to move or pull on their neighbours normally were chosen for further observation. All experiments were performed at 15 °C to slow cell movement to a rate of less than 10 μ m min⁻¹. As goldfish can survive at temperatures from 8 to 22 °C, 15 °C can be considered physiological.

Movement of actin filaments

To follow the dynamic behaviour of actin filaments, fluorescence was activated in a bar-shaped region in the lamellipodium (Fig. 4a-f). We then recorded images of the fluorescence distribution in the activated cell every 6 or 11 seconds, together with paired phase-contrast images which allowed us to determine the position of the cell margin. The activated bar generally remained fixed in space as the cell moved forward, and therefore moved backward towards the nucleus in the moving frame of reference of the cell. When the activated bar reached the rear area of the cell, the bar was sometimes seen to curve around the cell body (Fig. 4c, d). In all cells observed, the activated zone decreased

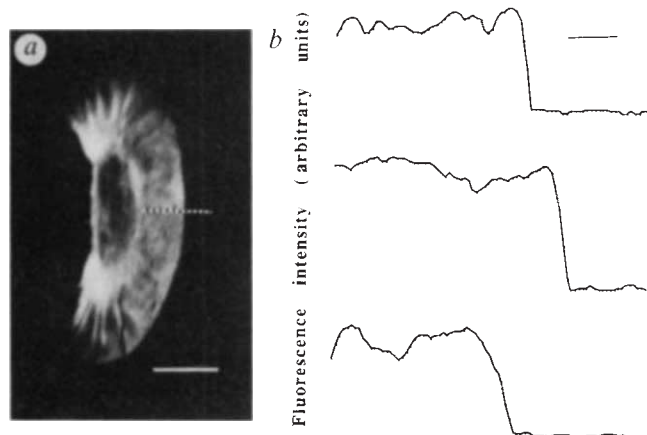


FIG. 3 Actin filament distribution is uniform in goldfish epithelial keratocytes. **a**, Typical keratocyte labelled with rhodamine-phalloidin. Scale bar, 10 μ m. **b**, Profile of fluorescence intensity in three representative cells. The profiles were measured through the lamellipodia, in positions corresponding to the dotted line in **a**. Scale bar, 2 μ m.

METHODS. Keratocytes were prepared from isolated scales of goldfish (*Carassius*)⁴¹ and cultured in Fish Ringer's solution⁴¹ supplemented with 20% Amphibian Culture Medium (Gibco). Fluorescence profiles were determined using Image 1 line-intensity scans.

in intensity over time, presumably due to filament turnover (see below). Fluorescence intensity profiles revealed that the bar did not spread over time, and that at least 90% of the activated actin seemed to behave as a coherent unit both in rapidly moving and in slowly moving cells (Fig. 4*i, j*).

To check that the persistent bar observed after activation in the lamellipodium was due to CR-actin incorporation into the actin cytoskeleton, carbethoxylated-CR-actin, which is strongly inhibited in its ability to polymerize (see above), was injected into keratocytes. This actin was then activated and observed. In contrast to native CR-actin, the fluorescent signal did not persist as a coherent bar, but spread throughout the cell in a few seconds (Fig. 4*g, h*).

The rates of rearward actin transport (relative to the substrate and relative to the cell margin) and forward cell movement were determined for 46 cells where the activated CR-actin bar could be followed for at least 30 s (Fig. 5). For cells moving at rates ranging from 0.001 to 0.13 $\mu\text{m s}^{-1}$ (0.06–8.0 $\mu\text{m min}^{-1}$), the rate of movement of the activated actin bar toward the nucleus with respect to the cell edge was roughly equal to the speed of the cell. That is, the activated bar remained fixed relative to the substrate as the cell moved forward over it, regardless of the rate of cell movement.

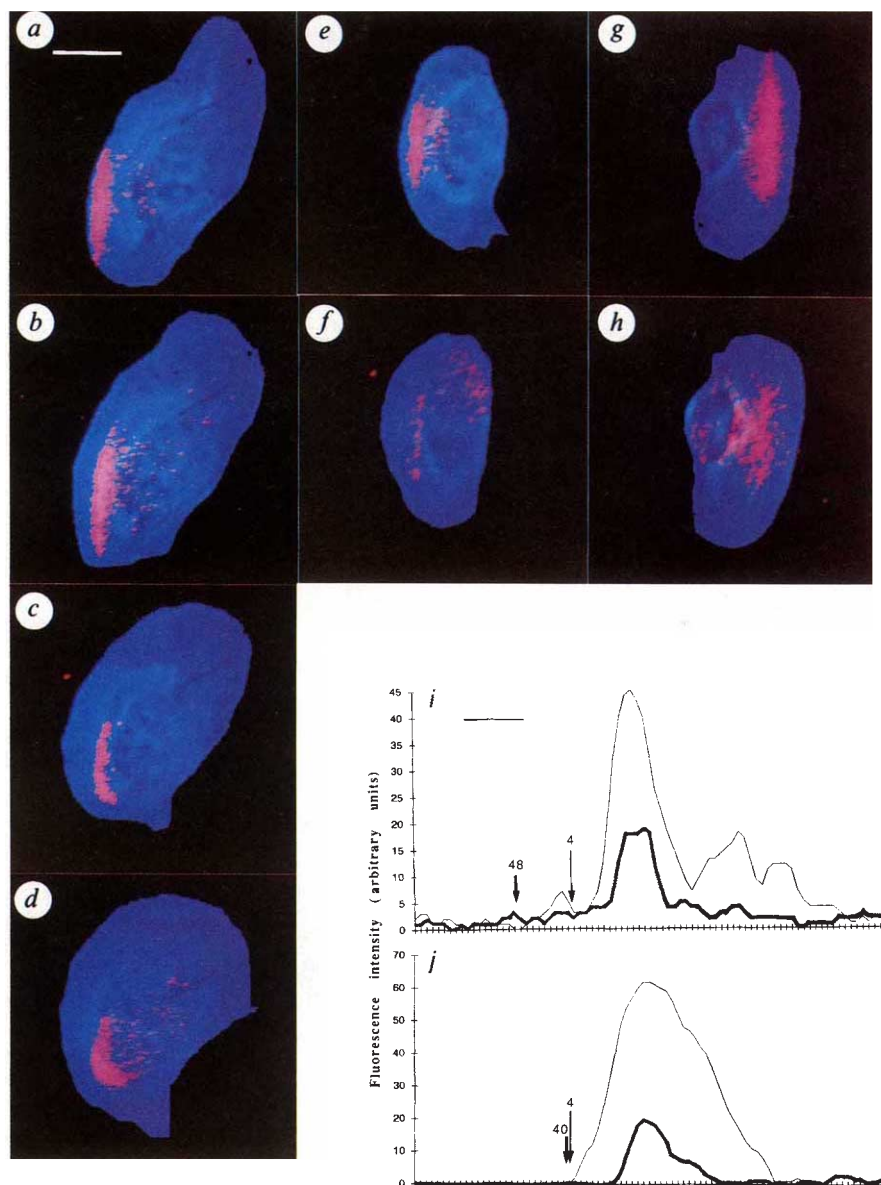
FIG. 4 Behaviour of actin microfilaments in moving keratocytes. Activated bar of CR-actin (pink) on a moving keratocyte (blue) 4 s (*a*), 48 s (*b*), 81 s (*c*) and 136 s (*d*) after activation. Scale bar, 10 μm . Note the curving of the activated bar in *c* and *d*. A second example 4 s (*e*) and 70 s (*f*) after activation. Activated region of carbethoxylated-CR-actin (not competent to polymerize) 1 s (*g*) and 6 s (*h*) after activation. Note that the activated bar is wider in *g* than in *a* or *e* due to rapid diffusion in 1 s. This diffusion is more evident in *h*. *i, j* Profiles of fluorescence intensity in a rapidly moving cell 4 s (thin line) and 48 s (thick line) after activation. The position of the cell margin at each time point is indicated by an arrow. Scale bar, 2 μm . *j*, Profiles of fluorescence intensity in a slowly moving cell 4 s (thin line) and 40 s (thick line) after activation. The position of the cell margin at each time point is indicated by an arrow. Note that the peak of fluorescence intensity remains stationary relative to the substrate in both cases. These profiles show fluorescence distribution in the lamellipodium only, as in Fig. 3. The decrease in fluorescence represents depolymerization and diffusion of activated subunits into the thick region of the cell near the nucleus, which can be seen in panels *e–f*.

METHODS. Keratocytes were injected and observed at 15 °C. Paired phase and fluorescence images were recorded <2 s apart, using an ISIT camera for both. Since no components of the optical system were moved and the wavelength of light in both cases was about the same (600–630 nm), the two images are exactly in register. Paired images were superimposed in pseudocolour using Image 1 software. The outline of the cell in the phase image was traced on the monitor and the rest of the image was blacked out so that the cell margin could be easily seen in the composite colour image. The bar in the cell shown in *a–d* persists an unusually long time, allowing visualization of the curving near the nucleus; the turnover in the cell shown in *e–f* is more typical. Photobleaching was measured by determining the total fluorescence intensity over time in cells in which the injected CR-actin had been completely activated. Bleaching was determined to be about 1.1 % per s of illumination (s.d.=0.3 %, $n=5$) which corresponds to 3.7 % per min of shuttered observation.

Ten of the 13 cells moving at rates of $<0.02 \mu\text{m s}^{-1}$ were either members of small clumps of cells¹¹ or bipolar cells with two lamellipodia pulling in opposite directions at each end¹². In both cases, opposing forces on the cell body prevented the cells from advancing rapidly. These cells appeared to remain healthy (they exhibited normal intracellular organelle motility) and their lamellipodia continued to generate force (the nuclei and cell bodies could be seen to be pulled by the opposing lamellipodia in a 'tug of war'). We were surprised that we did not observe any centripetal movement of actin filaments in these tethered cells, because such movement has been detected in nonmotile fibroblasts⁴ and growth cones⁵.

Filament turnover

As noted above, the fluorescence intensity in the activated bar in cells injected with CR-actin decreased rapidly. The turnover time of actin polymer in the keratocyte lamellipodium was determined from the rate of decrease in intensity. The average half-life of the signal arising from CR-actin in the activated region was 23 s (s.d.=6, $n=12$). This rapid turnover was not due to photobleaching of the fluorochrome; photobleaching over a 60-s period was <4% (see legend to Fig. 4). There was no correlation between the measured half-life and the rate of



cell movement or between the half-life and the position of the activated region in the lamellipodium.

Implications of filament movement

In the keratocyte lamellipodium, the actin filaments, once formed, remain fixed relative to the substrate. Furthermore, all of the filaments behave homogeneously as a single class; that is, we can detect no shear between different populations of microfilaments. Therefore, the cytoskeletal meshwork filling newly extended regions of the lamellipodium must arise solely by new actin polymerization and not by a sliding of preexisting filaments relative to each other^{21,22}. Similarly, the traction force which must be generated to drag the bulk of the cytoplasm forward also cannot involve shear between different populations of microfilaments²³.

The force responsible for cell locomotion is unknown. Candidates include force arising from the polymerization of actin itself, with its associated energy of ATP hydrolysis²⁴, turgor pressure resulting from local water influx²⁵, positive hydrostatic pressure resulting from a contraction of the rear of the cell²⁶ and force arising from molecular motors, such as myosin^{21,22,27}. We cannot yet distinguish among these possibilities, but note a few constraints imposed by our data. That we can detect no shear of actin filaments does not rule out a role for myosin in lamellipodial protrusion, but it does require that any myosin present be moving nonactin components on a unitary actin cytoskeleton rather than moving microfilaments relative to one

another. This activity is more likely to be performed by a single-headed myosin I than by conventional two-headed myosin II (ref. 27). This is in agreement with genetic data indicating that myosin II does not play a direct part in lamellipodial extension in *Dictyostelium*^{28,29}. We also note that as cells, whose membranes are tethered by interactions with other cells or by a second lamellipodium opposing the first, exhibit very slow rates of actin insertion at the leading edge, stretching of the membrane must be tightly coupled to new polymerization. This suggests at least a partial role for turgor and hydrostatic pressures in protrusion.

For the cell to move forward, it must generate traction relative to the substrate. Several classes of adhesion molecules interact indirectly with actin^{30,31}, and that the actin cytoskeleton in keratocytes is stationary relative to the substrate is consistent with the assumption that traction is generated by actin interacting with cell-surface adhesion proteins. The identity of the adhesive elements in keratocytes is unknown. Interference reflection microscopy of very similar cells from *Xenopus*³² indicates that keratocytes (like most other very rapidly locomoting cells, but unlike slower cells such as fibroblasts³³) do not make close focal contacts with their substrate.

In strong contrast to our results, centripetal movement of actin filaments in the absence of forward cell movement has been observed in fibroblasts⁴ and *Aplysia* growth cones⁵. Presumably in these cell types there must be some regulation of the adhesion molecule-cytoskeleton interaction. Keratocytes are highly specialized for motility and are not chemotactic towards any known agent. It seems possible that keratocytes may lack control mechanisms present in fibroblasts and growth cones which may trade efficiency in movement for the ability to respond to chemical signals. Polymorphonuclear leukocytes, though, are efficiently chemotactic and also move at top rates rivaling those of keratocytes³⁴. We need to measure actin movements in these other cell types as a function of cell speed to determine whether our general conclusion that lamellipodial protrusion is specifically correlated with new actin polymerization is true for cells other than keratocytes.

The dynamic behaviour of other elements of the lamellipodium is under investigation. In keratocytes¹³ as well as in polymorphonuclear leukocytes³⁴ the cell's plasma membrane moves forward passively with respect to the substrate (and therefore also with respect to the actin cytoskeleton) as the cell advances. By contrast, concavalin A-coated particles on cell surface glycoproteins may move rapidly forward or backward^{13,35}, in a manner distinct from the movements of either membrane or actin. We hope to investigate how these various components interact to produce at least four distinct behaviours in a single structure which is moving as a unit.

Implications of rapid turnover

It was unclear from previous work how the flux of actin subunits through the lamellipodium occurs. One possibility could be 'treadmilling' of individual long actin filaments⁴. According to this hypothesis (the treadmilling model), actin filaments are oriented with their barbed (active) ends towards the cell margin, and must be long enough to span the entire breadth of the lamellipodium. All polymerization is expected to occur at the cell margin, and all depolymerization at the rear of the lamellipodium. A second hypothesis (the nucleation-release model) does not require any particular orientation of actin filaments. Here a coherent flux of actin subunits through the lamellipodium occurs because of the movement of a meshwork of short, cross-linked actin filaments as a unit, with new filaments being created primarily at or near the leading edge and subsequently released. This model predicts that the average filament length be short compared with the dimensions of the lamellipodium. Polymerization and depolymerization may occur throughout the lamellipodium, as long as the polymerization rate is appropriately higher at the leading edge of moving cells.

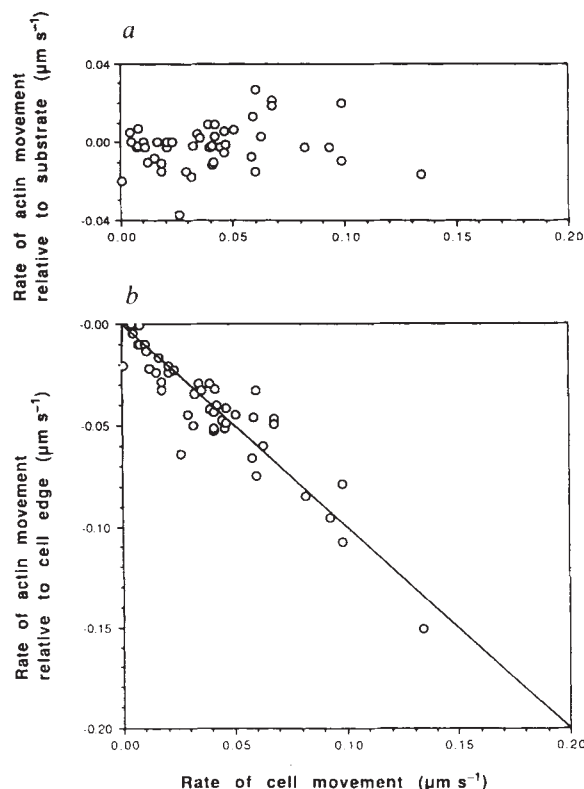


FIG. 5 Correlation between the rate of actin filament transport and the speed of cell locomotion. Plot of movement rate of activated bar of CR-actin relative to the substrate (a) or relative to the cell edge (b) as a function of cell speed. Each point represents an individual cell. Line shown in b is $x = -y$.

METHODS. The positions of the peak of actin fluorescence intensity and of the edge of the cell were recorded for each frame (every 6–11 s) over the period of observation (typically 30–90 s) using the Image 1 processor. Average rates for the period of observation were determined by fitting a straight line to each plot of position versus time.

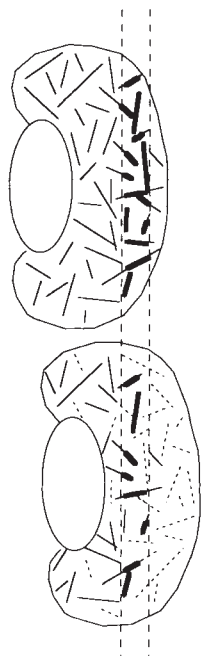


FIG. 6 Schematic representation of actin filament dynamics in the moving keratocyte according to the nucleation-release model. Top, heavy lines represent distribution of activated CR-actin immediately after activation; thin lines, unactivated polymer. Bottom, dashed lines represent newly formed filaments. The fluorescent signal decreases over time as filaments throughout the lamellipodium depolymerize and are replaced by new polymerization. The amount of new polymer present at the leading edge is intimately coordinated with the rate of cell movement. We have drawn the filaments as randomly oriented to contrast this model to the treadmilling model⁴, though in reality they may be at least partially oriented with their barbed ends towards the leading edge².

Our turnover results are most consistent with the nucleation-release model. We see a uniform rate of rapid turnover throughout the whole lamellipodium, with actin filaments depolymerizing just as rapidly near the periphery as near the nucleus. It is impossible to reconcile this observation with the treadmilling model without postulating exchange of actin subunits along the length of the filaments as well as at the ends. Such exchange is not observed in *in vitro* experiments. Also, if new polymerization occurred only at the leading edge, but depolymerization occurred uniformly over the lamellipodium, we would expect to see a steep gradient of actin filament density across the lamellipodium, with very little polymer remaining

near the cell body. A gradient of this type has been observed in the leading edge of fibroblasts³⁶, but we see no evidence for an actin filament density gradient in keratocytes (Fig. 3). We conclude that both polymerization and depolymerization of actin filaments must be occurring throughout the entire lamellipodium, to explain the uniform steady-state filament distribution. Thus the spatial pattern of filament turnover is consistent with the nucleation-release model, but inconsistent with the treadmilling model.

We also favour the nucleation-release model from kinetic arguments. The known rates of actin treadmilling and depolymerization are too slow for the treadmilling hypothesis to be viable. The rate of treadmilling of pure actin filaments *in vitro* is limited by the pointed (less active) end off-rate, and is about 100 times slower than would be required to account for the rates of cell movement we observe^{37,4}. The maximum rate of actin depolymerization observed *in vitro* from both pointed and barbed ends combined is sufficiently slow that the mean half-life of 23 s in keratocytes could only be observed for a population of actin filaments if their mean length were less than 0.5 μm ³⁷. Such short filaments in the lamellipodium are predicted by the nucleation-release model, but inconsistent with the treadmilling model. The true *in vivo* rates of actin treadmilling and depolymerization are unknown, so we cannot rule out the possibility that factors inside the cell increase the pointed end off-rate by 100 times, but believe that this is unlikely as no such activity has yet been purified.

To resolve this issue, it will be critical to have a detailed understanding of the organization of the keratocyte lamellipodium at the electron microscopic level. Available ultrastructural data on keratocytes does not seem to reveal a highly organized parallel meshwork in advancing lamellipodia¹¹. It would be most desirable to correlate the ultrastructure of the keratocyte lamellipodium with its dynamic behaviour, as has recently been done in macrophages³⁸. This work suggests that the turnover time of the lamellipodial cytoskeleton in macrophages is around 30 s, in good agreement with our measured value of 23 s in a different cell type using very different techniques.

A nucleation-release model for the dynamic organization of actin filaments in the keratocyte lamellipodium, derived from our data, is illustrated in Fig. 6. We have no information on how turnover occurs at the single filament level; depolymerization might occur from the pointed end (single-filament treadmilling), from the barbed end (dynamic instability) or both. We believe that the photoactivation system we have developed for observing actin filaments *in vivo* will be generally useful for determining actin behaviour and dynamics in many cell activities. □

Received 8 April; accepted 31 May 1991.

- Pollard, T. D. *Curr. Opin. Cell Biol.* **2**, 33–40 (1990).
- Small, J. V. *Electron Microsc. Rev.* **1**, 155–174 (1982).
- Yin, H. L. & Hartwig, J. H. *J. Cell Sci. Suppl.* **9**, 169–184 (1988).
- Wang, Y.-L. *J. Cell Biol.* **101**, 597–602 (1985).
- Forscher, P. & Smith, S. J. *J. Cell Biol.* **107**, 1505–1516 (1988).
- Fisher, G. W., Conrad, P. A., DeBiasio, R. L. & Taylor, D. L. *Cell Motil. Cytoskel.* **11**, 235–247 (1988).
- Svitkina, T. M., Neyfakh, A. A. Jr & Bershadsky, A. D. *J. Cell Sci.* **82**, 235–248 (1986).
- Mitchison, T. J. *J. Cell Biol.* **109**, 637–652 (1989).
- Sawin, K. E. & Mitchison, T. J. *J. Cell Biol.* **112**, 941–954 (1991).
- Euteneuer, U. & Schliwa, M. *Nature* **310**, 58–61 (1984).
- Kolega, J. *J. Cell Biol.* **102**, 1400–1411 (1986).
- Cooper, M. S. & Schliwa, M. *J. Cell Biol.* **102**, 1384–1399 (1986).
- Kucik, D. F., Elson, E. L. & Sheetz, M. P. *J. Cell Biol.* **111**, 1617–1622 (1990).
- Haugland, R. P. *Molecular Probes Handbook of Fluorescent Probes and Research Chemicals*. (Molecular Probes, Eugene, Oregon, 1989).
- Walker, J. W., Reid, G. P. & Trentham, D. R. *Meth. Enzym.* **172**, 288–301 (1989).
- dos Remedios, C. G., Miki, M. & Barden, J. A. *J. Muscle Res. Cell Motil.* **8**, 97–117 (1987).
- Wang, Y.-L. & Taylor, D. L. *J. Histochem. Cytochem.* **28**, 1198–1206 (1980).
- Hegy, G., Premecz, G., Sain, B. & Muhlrad, A. *Eur. J. Biochem.* **44**, 7–12 (1974).
- Amato, P. A. & Taylor, D. L. *J. Cell Biol.* **102**, 1074–1084 (1986).
- Kreis, T. E., Geiger, B. & Schlessinger, J. *Cell* **29**, 835–845 (1982).
- Smith, S. J. *Science* **242**, 708–715 (1988).
- Egelhoff, T. T. & Spudich, J. A. *Trends Genet.* **7**, 161–166 (1991).
- Huxley, H. E. *Nature* **243**, 445–449 (1973).

- Hill, T. L. & Kirschner, M. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 490–494 (1982).
- Oster, G. F. & Perelson, A. S. *J. Cell Sci. Suppl.* **8**, 35–54 (1987).
- Taylor, D. L. & Condeelis, J. S. *Int. Rev. Cytol.* **56**, 57–144 (1979).
- Korn, E. D. & Hammer, J. A. III *Curr. Opin. Cell Biol.* **2**, 57–61 (1990).
- Knecht, D. A. & Loomis, W. F. *Science* **236**, 1081–1085 (1987).
- DeLozanne, A. & Spudich, J. A. *Science* **236**, 1086–1091 (1987).
- Burridge, K., Fath, K., Kelly, T., Nuckolls, G. & Turner, C. A. *Rev. Cell Biol.* **4**, 487–525 (1988).
- Takeichi, M. *Development* **102**, 639–655 (1988).
- Bereiter-Hahn, J. & Voith, M. *Cell Motil. Cytoskel.* **10**, 528–536 (1988).
- Trinkaus, J. P. *Cells into Organs: The Forces that Shape the Embryo 202* (Prentice-Hall, Englewood Cliffs, New Jersey 1984).
- Lee, J., Gustafsson, M., Magnusson, K.-E. & Jacobson, K. *Science* **247**, 1229–1233 (1990).
- Kucik, D. F., Elson, E. L. & Sheetz, M. P. *Nature* **340**, 315–317 (1989).
- Symons, M. & Mitchison, T. J. *J. Cell Biol.* (in the press).
- Pollard, T. D. & Cooper, J. A. *Rev. Biochem.* **55**, 987–1035 (1986).
- Rinnerthaler, G., Herzog, M., Klappacher, M., Kunka, H. & Small, J. V. *J. struct. Biol.* **106**, 1–16 (1991).
- Pardee, J. D. & Spudich, J. A. *Meth. Enzym.* **85B**, 164–181 (1982).
- Rinnerthaler, G., Geiger, B. & Small, J. V. *J. Cell Biol.* **106**, 747–760 (1988).
- Cooper, M. S. & Schliwa, M. *J. Neurosci. Res.* **13**, 223–244 (1985).

ACKNOWLEDGEMENTS. We acknowledge the contribution of M. Symons to this work, and thank C. Field, B. Alberts, L. Belmont, K. Sawin and T. Hyman, for discussions and comments on the manuscript. This work was supported by the Packard and Searle Foundations (T.J.M.) and the Markey Foundation through the UCSF microscopy facility. J.A.T. is a Howard Hughes Medical Institute Pre-doctoral Fellow.

An unusually strong Einstein ring in the radio source PKS1830–211

D. L. Jauncey*, J. E. Reynolds†, A. K. Tzioumis‡, T. W. B. Muxlow‡, R. A. Perley§, D. W. Murphy||, R. A. Preston||, E. A. King¶, A. R. Patnaik‡, D. L. Jones||, D. L. Meier||, D. J. Bird#, D. G. Blair**, J. D. Bunton††, R. W. Clay#, M. E. Costa**, R. A. Duncan*, R. H. Ferris*, R. G. Gough*, P. A. Hamilton¶, D. W. Hoard||, A. Kembell‡‡, M. J. Kesteven*, E. T. Lobdell||, A. N. Luiten**, P. M. McCulloch¶, J. D. Murray*, G. D. Nicolson‡‡, A. P. Rao§§, A. Savage##, M. W. Sinclair*, L. Skjerve||, L. Taaffe#, R. M. Wark* & G. L. White*

* Australia Telescope National Facility, CSIRO, Epping, New South Wales 2121, Australia

† Mount Stromlo and Siding Spring Observatories, Canberra, Australian Capital Territory 2611, Australia

‡ Nuffield Radio Astronomy Laboratory, Jodrell Bank, Macclesfield, Cheshire, SK11 9DL, UK

§ National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA

|| Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

¶ Physics Department, University of Tasmania, Hobart, Tasmania 7001, Australia

Department of Physics, University of Adelaide, Adelaide, South Australia 5001, Australia

** Department of Physics, University of Western Australia, Nedlands, Western Australia 6009, Australia

†† Division of Radiophysics, CSIRO, Epping, New South Wales 2121, Australia

‡‡ Hartebeesthoek Radio Astronomy Observatory, Johannesburg, South Africa

§§ Radio Astronomy Centre, Tata Institute of Fundamental Research, Udhagamandalam 643 001, India

UK Schmidt Telescope Unit of the Anglo-Australian Observatory, Coonabarabran, New South Wales 2857, Australia

RADIO observations of the strong, flat-spectrum radio source PKS1830–211 revealed a double structure, with a separation of 1 arcsec, suggesting that it might be a gravitationally lensed object¹. We have now obtained high-resolution radio images of PKS1830–211 from several interferometric radiotelescope networks, which show an unusual elliptical ring-like structure connecting the two brighter components. The presence of the ring, and the similarity of the two brighter spots, argue strongly that this is indeed a gravitationally lensed system, specifically an Einstein ring in which lens and lensed object are closely aligned. Although the source is close to the galactic plane, it seems that both the lens and background (lensed) object are extragalactic. This object is one hundred times brighter than either of the two previously discovered radio Einstein rings, and is among the six brightest flat-spectrum sources in the sky. Its brightness makes it a peculiar object: it must involve either a chance alignment of a lensing object with an unusually bright background source, or an alignment with a less bright object but amplified to an unusual degree.

In November 1988, we used very long baseline interferometry (VLBI) to obtain images of PKS1830–211 at 2.3 GHz with seven Southern Hemisphere radiotelescopes including the Australia Telescope Compact Array. These showed that PKS1830–211 has two compact components which have a separation of 1 arcsec, but which account for less than half the total flux density of 11.7 Jy. The total flux had, however, increased by 25% over

the earlier Parkes values. Observations with MERLIN² followed at 1.7 GHz in June 1989 and the image at 0.25-arcsec resolution, Fig. 1, revealed a complete elliptical ring of low surface brightness with the two strong compact VLBI components at either end of the major axis. The ring has a deep central minimum (less than 0.3% of the peak) and striking rotational symmetry.

A subsequent 8.4-GHz VLA image with the same resolution (Fig. 2) confirms the structure found with MERLIN. Comparison of the two images shows that both compact components possess flat radio spectra and that the low-surface-brightness ring has a steep spectrum which steepens further with increasing anticlockwise separation from the compact components. Significant linear polarization of a few per cent was also detected. Our 2.3-GHz Southern Hemisphere VLBI experiment (SHEVE³) image (Fig. 3) shows only the structure of the compact components, as the ring is completely resolved on the 275-km minimum baseline. The two components have similar angular sizes of 15×30 mas, are separated along position angle 44° and have peak brightness temperatures of $\sim 10^9$ K.

These properties of PKS1830–211 lead us to the conclusion that it is a gravitational lens system—specifically, an Einstein ring—formed by the gravitational imaging of a background radio source by a foreground mass concentration⁴. Other known astrophysical systems do not produce the same combination of ring-like morphology with such a deep central null, nonthermal radio spectrum, complex spectral index distribution, linear polarization, variability and high brightness temperature. Among radio sources, only two other possible Einstein ring systems are currently known, MG1131+0456 and MG1654+1346 (refs 5, 6). The similarity in the overall morphology between

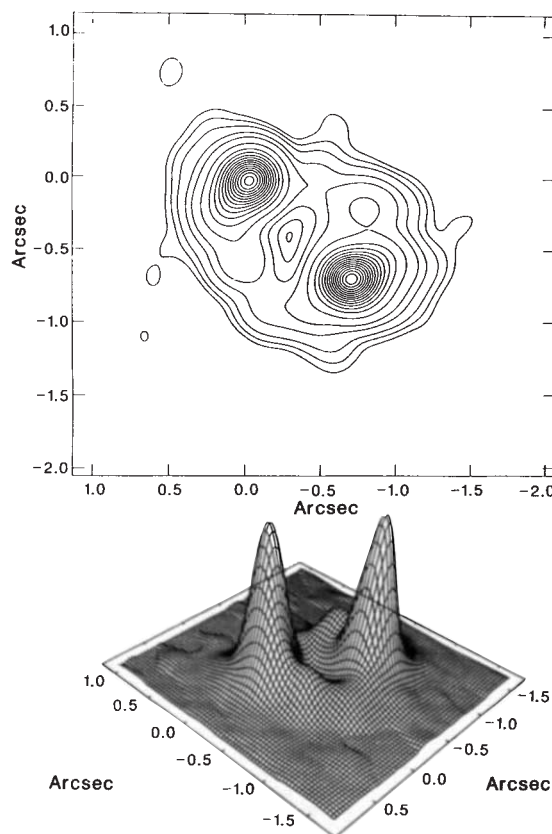


FIG. 1 The MERLIN 1.692-GHz hybrid map of PKS1830–211 shown as both a contour map and in relief. The restoring beam is a circular gaussian with a full width at half maximum (FWHM) of $0.25''$. Contour levels are at $-0.005, 0.005, 0.01, 0.02, 0.05, 0.1, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25$ and 3.5 Jy per beam. The relief image plots the square root of the amplitude to emphasize the presence of the ring-like structure.

these objects and PKS1830–211 is remarkable, although PKS1830–211 is more than two orders of magnitude stronger.

The observed morphology can readily be interpreted as the double image of the flat-spectrum core of the background source, whereas the steep-spectrum jet seems to be more closely aligned along the line of sight to the lensing object and is bent to form the ring-like structure in Figs 1 and 2. There is no evidence for a third compact component above the noise level in either the SHEVE, MERLIN or VLA images. The source PKS1830–211 is so strong that the odd-numbered (third) image predicted by theory⁷ may eventually be found, whereas none has been conclusively detected in the other weaker radio lenses.

We have examined the UK Schmidt plates for an optical identification, and although the field is crowded ($l = 12.2^\circ$, $b = -5.7^\circ$) no object is found above the plate limit at the radio position on the J-band or R-band plates. An object of magnitude 20, close to the radio centroid position¹, RA 18 h 30 min 40.62 s, $-21^\circ 06' 00.3''$ (B1950), appears on both the Schmidt I plate and on an I-band CCD image taken with the Anglo-Australian Telescope⁸. There is no indication of anything unusual near the radio position, no obvious H II region, planetary nebula, bright star, galaxy or globular cluster. The expected extinction is ~ 2.4 magnitudes in the V band (ref. 9), or ~ 1 magnitude in the I band. No object has been found in the IRAS Point Source or Einstein IPC catalogues at the position of PKS1830–211.

As PKS1830–211 lies close to the direction of the Galactic Centre, it might be expected that the gravitational lens lies in our Galaxy. This cannot be the case. For a point-mass deflector, the ratio of lens mass, M , to the lens distance, D_l , is given by

$$M/D_l = (\Phi/2)^2 (D_s/D_{sl}) c^2 / 4G,$$

where Φ is the angular separation of the images and D_s and D_{sl} are the distances from the observer to the source and from the source to the lens respectively. For $\Phi = 1''$, this ratio is $30 M_\odot$ per parsec or greater, which puts any stellar gravitational lens of mass up to $150 M_\odot$ within 5 pc. The absence of anything brighter than magnitude 20 on the UK Schmidt plates clearly precludes a galactic star lens unless an unreasonably large mass-to-light ratio is assumed. Moreover, proper motions and galactic rotation imply lifetimes of less than a few years for a galactic lens; PKS1830–211 has been a radio source of ~ 10 Jy for at least 25 years.

At kiloparsec distances where the lensing object could be

obscured, a globular cluster with mass of $10^5 M_\odot$ is necessary to produce an image separation of $1''$. Typical globular clusters have linear sizes of ~ 10 pc, which at kiloparsec distances imply arcminute, not arcsecond, sizes. In addition, such a globular cluster should be visible on the Schmidt plates, as the expected extinction is modest.

Such ring-like structure in a radio source at low galactic latitude might indicate that PKS1830–211 is a H II region, planetary nebula or supernova remnant. The brightness temperature of $\sim 10^9$ K and variability of the compact components, and the presence of linear polarization, rule out a H II region or planetary nebula origin. There is also strong evidence against the supernova remnant origin: VLBI observations taken six months apart on the Tidbinbilla–Hobart baseline constrain the rate of separation of the two components to -2^{+3}_{-2} mas yr⁻¹, or a velocity of -100^{+150}_{-100} km s⁻¹ for a distance of 10 kpc. This is an order of magnitude less than would be expected from such a young, $1''$ diameter, supernova remnant.

A black hole of mass $3 \times 10^5 M_\odot$ near the Galactic Centre could act as a lens, but the precise angular alignment required is most unlikely. For example, an overall amplification factor of 10 in a lensed ring of $1''$ diameter requires an alignment between the background source and the lensing object of $< 0.050''$ (see, for example, equation 18 of ref. 10). As the background density of radio sources with flux densities above 1 Jy is ~ 1 per 30 square degrees¹¹, the *a priori* probability of a sufficiently accurate alignment is only $\sim 2 \times 10^{-11}$ per lensing object. Even for a population of 10^6 black holes, the equivalent in mass to the entire visible galaxy, the likelihood of such an event is remote. The overall probability does not depend strongly on the assumed amplification factor, as the increasing number of background sources at lower flux densities is balanced by the tighter angular alignment required to maintain the observed 10 Jy. It follows that both the lens and the source producing the strong ratio image are almost certainly extragalactic.

Even without redshifts for the object and lens, it is still possible to derive order-of-magnitude estimates of some of the lens properties. Press and Gunn¹² report that the scattering mass is weakly dependent on source redshift and, for a separation of $1''$, the likely lensing mass is around $2 \times 10^{11} M_\odot$, similar to that found for the lensing galaxy in 1654 + 1346 (ref. 13). In addition,

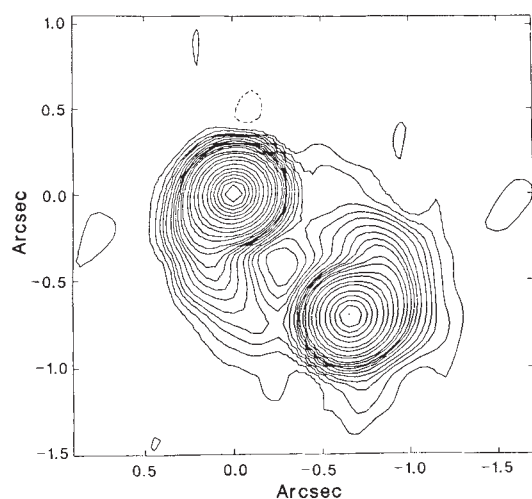


FIG. 2 Contour plot of the 8.4-GHz VLA radio image of PKS1830–211. The restoring beam is a circular gaussian with FWHM of $0.25''$. Contour levels are at $-0.15, 0.15, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 7.5, 10, 15, 20, 30, 40, 50, 60, 70, 80$, and 90% of the peak flux density (3.47 Jy per beam).

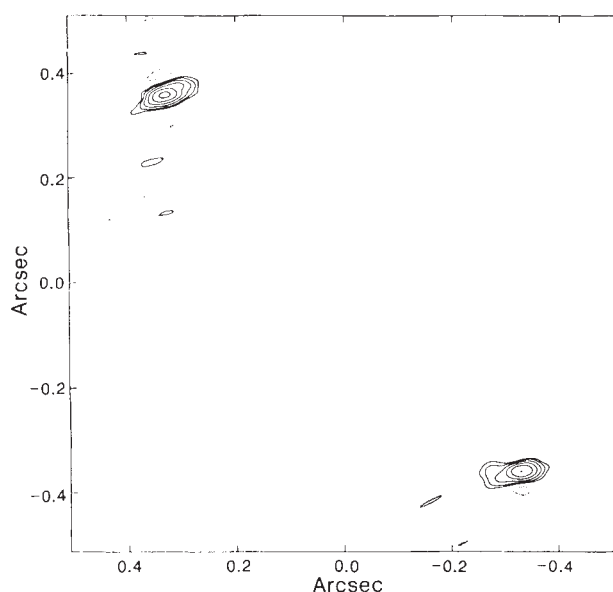


FIG. 3 Contour plots of the 2.3-GHz SHEVE image of PKS1830–211. The restoring beam is an elliptical gaussian with FWHM of 30×10 milliarcseconds in position angle -80° . Contour levels are at $-2.5, 2.5, 5.0, 10, 20, 40$ and 80% of the peak flux density (1.41 Jy per beam).

the expected differential time delay between the two compact image paths is $\sim 5 \times 10^6$ seconds (about two months). Continued monitoring of the flux density and structure of PKS1830–211 is clearly necessary on timescales shorter than this. Given the small angular separation of the radio components and the crowded nature of the field, Hubble Space Telescope observations may be necessary to detect, resolve and obtain redshifts for any lensed optical images or the lensing galaxy. Infrared observations may also help to penetrate any obscuration so close to the galactic plane.

As PKS1830–211 is known to be variable, continued monitoring of the two components may yield considerable information on the path difference between the two compact images and ultimately lead to a determination of the Hubble constant H_0 , given redshifts for the object and lens. Detailed modelling of the source must await optical identification and redshift measurements for the object and for the lensing galaxy. Given the similar morphologies of PKS1830–211, MG1131+0456 and MG1654+1346, the success in modelling MG1131+0456 and MG1654+1346 (refs 13, 14) may be expected to be repeated for PKS1830–211.

The presence of the two strong lensed milliarcsecond variable cores in PKS1830–211 will allow several sensitive tests to be performed. In particular, VLBI images may reveal time-dependent changes in structure. Most variable active galactic nuclei show superluminal expansion¹⁵, but any such expansion in PKS1830–211 will be amplified both in size and intensity by the lensing galaxy. In this case, measurements of the apparent internal expansion within each compact component may give an estimate of this amplification factor. Additionally, differences in the component angular sizes may give an independent measure of the time delay in the two light paths.

The intensity of PKS1830–211 is a puzzle, as the all-sky catalogue of Wall and Peacock¹⁶ lists only five flat-spectrum extragalactic sources at 2.7 GHz stronger than PKS1830–211, and lensing occurs only once per few thousand radio sources¹⁷. It is one hundred times stronger at radio wavelengths than the other two candidates for Einstein rings and raises the question of whether other strong gravitational lens or Einstein ring radio sources may be found. Is PKS1830–211 merely a particularly unusual chance alignment with a large amplification, or are there indeed more strong Einstein rings and gravitational lenses to be found? If so, where are they and how are they to be recognized? □

Role of heterogeneous conversion of N_2O_5 on sulphate aerosols in global ozone losses

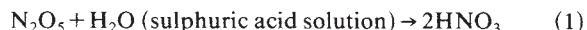
Jose M. Rodriguez, Malcolm K. W. Ko & Nien Dak Sze

Atmospheric and Environmental Research, Inc., Cambridge, Massachusetts 02139, USA

THERE is a serious discrepancy between estimates of the downward trend in the column abundance of ozone at middle and high latitudes derived from ground-based and satellite data, and those obtained by models that calculate the effect of increases in atmospheric chlorine concentrations, but include only gas-phase chemistry^{1,2}. Recent measurements^{3,4} of the reaction rate of N_2O_5 on sulphate aerosols yield very fast rates for a wide range of water content, indicating that this reaction could take place in the stratospheric sulphate aerosol layer, which is present around the globe and year-round between ~ 14 –25 km altitude. Here we include this reaction in a two-dimensional model and find that the calculated decadal ozone trends at both high and middle latitudes agree much more closely with the trends deduced from observations. Inclusion of the reaction also significantly increases the predicted concentrations of key species such as OH, ClO and HNO_3 , and decreases those of NO and NO_2 . Measurements of these species in and near the sulphate layer are needed to confirm the importance of this reaction in the observed decrease of atmospheric ozone.

Previous calculations^{5–7} indicated that the observed ozone trends at northern high latitudes in winter could be a result of heterogeneous chemistry occurring in arctic polar stratospheric clouds (PSCs). Ozone at lower latitudes could also be affected by processes occurring outside the polar vortices and extending beyond the winter season. Some models predict that fast heterogeneous reactions of ClONO_2 with either H_2O or HCl on sulphate aerosols could induce large ozone depletions⁸, particularly after volcanic eruptions^{9,10}. Laboratory kinetic data^{11,12} and observations¹³ indicate, however, that the rates of these reactions may be extremely slow at temperatures and water contents typical of the sulphate layer, raising the question of whether these rates could have a significant global impact.

Recent laboratory measurements^{3,4} indicate a surprisingly fast rate for the heterogeneous reaction



with reaction efficiencies ranging from ~ 0.06 to 0.1 for a wide range of water content in the reacting surface. This reaction could therefore occur in the global sulphate aerosol layer at all latitudes and seasons, and may represent an important sink for NO_x ($\text{NO} + \text{NO}_2 + \text{NO}_3 + 2 \times \text{N}_2\text{O}_5$). Furthermore, reaction (1) followed by photolysis of HNO_3 constitutes a considerable stratospheric source of reactive hydrogen radicals (HO_x).

Although the impact of this mechanism should ideally be assessed by using a three-dimensional model, a two-dimensional approach⁵ provides an efficient tool for assessing the effect of large-scale, long-term mechanisms. We have incorporated reaction (1) in our two-dimensional model. No other heterogeneous mechanisms have been included in this study. The rate is represented by the product of the gas–aerosol collision frequency and a reaction efficiency which is assumed to be the same at all latitudes and seasons. We also assume that the stratospheric aerosol layer is saturated and all the HNO_3 produced by reaction (1) is immediately released into the gas phase. Values for the aerosol collision frequencies at mid-latitudes have been taken from ref. 14. These values are about twice as high as the ‘background’ 1979 values over Laramie, Wyoming⁹. We assume that the aerosol loading does not vary from year to year, although recent results¹⁵ indicate that the aerosol abundances in the

Received 16 August 1990; accepted 24 May 1991.

1. Rao, A. P. & Subrahmanyam, R. *Mon. Not. R. astr. Soc.* **231**, 229–236 (1988).
2. Thomasson, P. Q. *J. R. astr. Soc.* **27**, 413–431 (1986).
3. Jauncey, D. L. et al. in *Frontiers of VLBI, Proc. Int. Symp. on VLBI Space Observatory Programme and Proc. Millimetre-Wave VLBI Workshop* (eds Hirabayashi, H., Inoue, M. & Kobayashi, H.) (Universal Academy Press, Tokyo, 1990).
4. Einstein, A. *Science* **84**, 506 (1936).
5. Hewitt, J. N. et al. *Nature* **333**, 537–540 (1988).
6. Langston, G. I. et al. *Astr. J.* **97**, 1283–1290 (1989).
7. Blandford, R. D. & Kochanek, C. S. *Astrophys. J.* **321**, 658–675 (1987).
8. Subrahmanyam, R., Narasimha, D., Rao, A. P. & Swarup, G. *Mon. Not. R. astr. Soc.* **246**, 263–272 (1990).
9. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. G. *Second Reference Catalogue of Bright Galaxies*, **32** (University of Texas Press, 1976).
10. Refsdal, S. *Mon. Not. R. astr. Soc.* **128**, 295–306 (1964).
11. Wall, J. V., Shimmins, A. J. & Merkleij, J. K. *Aust. J. Phys. Astrophys. Suppl.* no. 19 (1971).
12. Press, W. H. & Gunn, J. E. *Astrophys. J.* **185**, 397–412 (1973).
13. Langston, G. I., Conner, S. R., Lehar, J., Burke, B. F. & Weiler, K. W. *Nature* **344**, 43–45 (1990).
14. Kochanek, C. S., Blandford, R. D., Lawrence, C. R. & Narayan, R. *Mon. Not. R. astr. Soc.* **238**, 43–56 (1989).
15. Zensus, J. A. & Pearson, T. J. *IAU Symp.* no. 129: *The Impact of VLBI on Astrophysics and Geophysics* (eds Reid, M. J. & Moran, J. M.) 7–16 (Kluwer, Dordrecht, 1988).
16. Wall, J. V. & Peacock, J. A. *Mon. Not. R. astr. Soc.* **216**, 173–192 (1985).
17. Hewitt, J. N. et al. *Lecture Notes in Physics 330: Gravitational Lenses* (eds Moran, J. M., Hewitt, J. N. & Lo, K. Y.) 147–157 (Springer, Berlin, 1989).

ACKNOWLEDGEMENTS. We thank the Canberra Deep Space Communications Complex and the European Space Agency for use of their facilities at Tidbinbilla and Gungahlin. E.A.K., D.J.B. and G.L.W. acknowledge National Research Fellowships. This research was also supported by CSIRO-Universities grants. Part of this research was carried out at the Jet Propulsion Laboratory of the California Institute of Technology, under contract to NASA. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under contract to the NSF. The Australia Telescope National Facility is operated in association with the Division of Radiophysics by CSIRO.

TABLE 1 Comparison of calculated and observed trends in total ozone concentration (% per decade)

Season	Latitude	Analysis of Dobson data				Model results (1969–85)			
		OTP† (69–86)	WMO (70–88)§	WMO (78–88)	Bojkov <i>et al.</i> ²⁶ (58–86)	$\gamma_1=0.0\parallel$	$\gamma_1=0.01$	$\gamma_1=0.06$	$\gamma_1=0.1$
Winter*	35° N	–1.4 (0.8)‡	–1.8 (0.5)	–3.3 (1.0)	–1.2 (0.4)	–1.7	–2.0	–2.5	–2.8
	45° N	–2.8 (0.9)	–2.3 (0.5)	–2.7 (0.7)	–2.1 (0.3)	–1.9	–2.2	–2.9	–3.3
	55° N	–3.7 (0.9)	–2.7 (0.6)	–2.2 (1.1)	–3.0 (0.5)	–2.1	–2.5	–3.3	–3.7
Summer†	35° N	–1.1 (0.5)	–0.5 (0.5)	–0.8 (0.7)	–0.7 (0.3)	–1.3	–1.7	–2.2	–2.5
	45° N	–1.1 (0.4)	–0.3 (0.5)	–1.6 (0.5)	–0.7 (0.3)	–1.3	–1.7	–2.3	–2.6
	55° N	–0.12 (0.5)	–0.2 (0.6)	–2.4 (0.7)	–0.7 (0.3)	–1.3	–1.8	–2.5	–2.8

* December–March inclusive.

† May–August inclusive.

‡ NASA Ozone Trends panel¹; numbers in parentheses refer to the years covered by data.§ WMO/UNEP report²⁵, table 2.2.5. Analysis in this report extended OTP results by using different statistical techniques and extending data base.|| WMO/NASA report²⁵, see discussion of their Table 2.2.8 for details. $\parallel$ γ_1 denotes adopted reaction efficiency for $\text{N}_2\text{O}_5 + \text{H}_2\text{O}$.# Numbers in parentheses denote the standard error (2σ).

unperturbed stratosphere could have increased by $\sim 5\%$ per year during the last decade (50% over the 10 years). We estimate that the uncertainty in the aerosol collision frequencies that we have adopted is a factor of 0.5 to 2. As N_2O_5 abundances increase at night and at high winter latitudes as solar dissociating radiation decreases, assessment of the impact of reaction (1) requires detailed accounting of the diurnal variation of all the species. Here we explicitly calculate the diurnal behaviour of reactive species.

We have performed calculations for stratospheric chlorine contents ranging from 1.5 to 4.6 parts per 10^9 by volume (p.p.b.v.). We estimate that chlorine loadings of 1.5, 2.4, 2.9 and 3.4 p.p.b.v. correspond to conditions in 1969, 1979, 1985 and 1990, respectively. Extrapolations into the future adopting the World Meteorological Organization/United Nations Environment Program⁵ scenario B (50% cutback in the Montreal prod-

ucts by the year 2000) yield chlorine loadings of 4.6 p.p.b.v. by the year 2010¹⁶. We also include the effect of increasing tropospheric abundances of CH_4 (1% per year) and N_2O (0.25% per year) in our simulations. The simulations do not include the variability of the solar flux during the solar cycle, nor any interannual variations in temperature and circulation.

The additional OH source and NO_x sink introduced by reaction (1) increase the calculated noon concentrations of OH by 20–40% at high latitudes in winter, and reduces the calculated NO_x by 60–70% at high latitudes in winter and by 30–50% at mid-latitudes in the vicinity of the sulphate layer. These changes also affect other key trace species, particularly in the chlorine family¹⁷. Concentrations of HCl are shifted to reactive chlorine ($\text{ClO}_x = \text{Cl} + \text{ClO} + 2 \times \text{Cl}_2\text{O}_2 + \text{OCIO} + \text{ClNO}_3$) by faster conversion of HCl through reaction with OH. Reductions in NO slow down regeneration of chlorine through the reaction of ClO

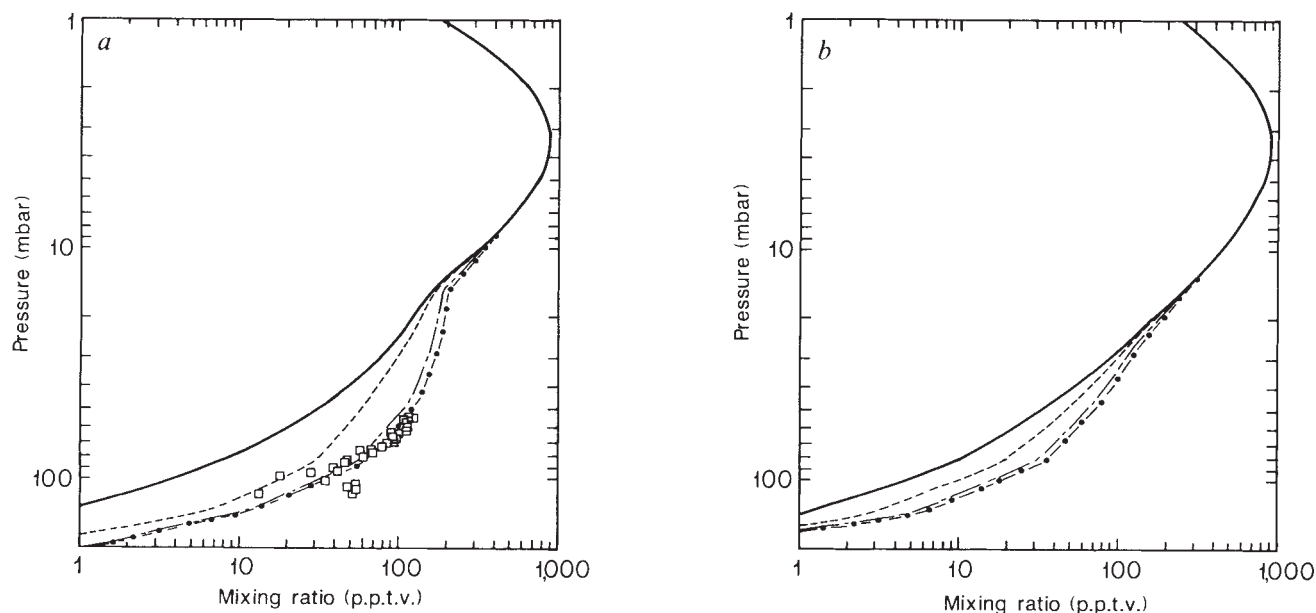


FIG. 1 Calculated mixing ratios of ClO (in parts per 10^{12}) at noon for 1988 conditions at 57° N, for *a*, winter (28 February 1988) and *b*, summer (31 August 1988). Calculations assume no heterogeneous chemistry (solid line)

and efficiencies for reaction (1) equal to 0.01 (----), 0.06 (— · —) and 0.1 (— ● —). The recalibrated measurements of ref. 21, obtained at 60° N on 13 February, are shown for comparison in *a* (squares).

with NO. As a result, the time constant for reconverting ClO_x to HCl through the reaction of chlorine with CH_4 is longer, and high concentrations of ClO_x can be maintained. Finally, reductions in NO_2 shift the partitioning between ClO and ClONO_2 in favour of ClO, leading to enhanced concentrations of ClO in the lower stratosphere and potentially larger negative ozone trends.

The impact on calculated concentrations of trace species should allow us to test the occurrence of reaction (1) in the stratosphere. The abundance of HNO_3 calculated from reaction (1) increases by 30–40% in the sulphate layer at high winter latitudes. As noted by other authors^{9,18–20}, these higher values agree better with observations of HNO_3 at these locations and season.

The impact of reaction (1) on ClO is illustrated in Fig. 1 for 57° N in winter and summer. Although the signature of reaction (1) is more pronounced during winter, enhancement in ClO by factors of 3–4 is still calculated during summer. Enhancements by a factor of two (not shown) are calculated even at mid-latitudes during winter. Recalibrated measurements (W. Brune, personal communication) from ref. 21 are compared with calculations in Fig. 1a. The good agreement may be fortuitous, given the difficulties in comparing results from a zonally-averaged two-dimensional model with a single aircraft profile, which may include the effects of exported vortex air or local downward motions not included in the model. The evidence from existing ClO measurements at lower latitudes is inconclusive. Enhanced

concentrations of ClO have been observed²² as far south as 38° N at altitudes between 14 and 20 km during February 1989. Other measurements²³ of ClO at mid-latitudes during summer, however, do not seem to show a signature of heterogeneous chemistry⁸.

Our calculations also indicate that reaction (1) decreases the calculated ratio of NO to total inorganic nitrogen (NO_x) by factors of two to three in or near the sulphate layer, even at mid-latitudes and during summer. Coincident measurements of all species in the NO_x family are available from the ATMOS experiment for 3–4 days in 1985²⁴, and the technology is available for making similar measurements aboard balloons and the ER-2. Uncertainties in the ATMOS measurements in the lower stratosphere, and in the possible effect of azonal motions of the trace gases on the NO/NO_x ratio, preclude unequivocal identification of heterogeneous chemistry from the few profiles taken by ATMOS. In summary, preliminary analysis of existing observations of ClO and NO_x does not either firmly establish or rule out the occurrence of reaction (1). More detailed analysis, including the dynamical conditions for the specific day and location of the measurements, and a larger data base should help in determining the range of locations and seasons in which reaction (1) occurs.

The calculated ozone trends for 1979–90 and for 1990–2010 are shown in Fig. 2 for cases with no heterogeneous chemistry (2a and c), and when reaction (1) occurs with an efficiency of 0.1 (2b and d). Trends in column ozone calculated with reactions

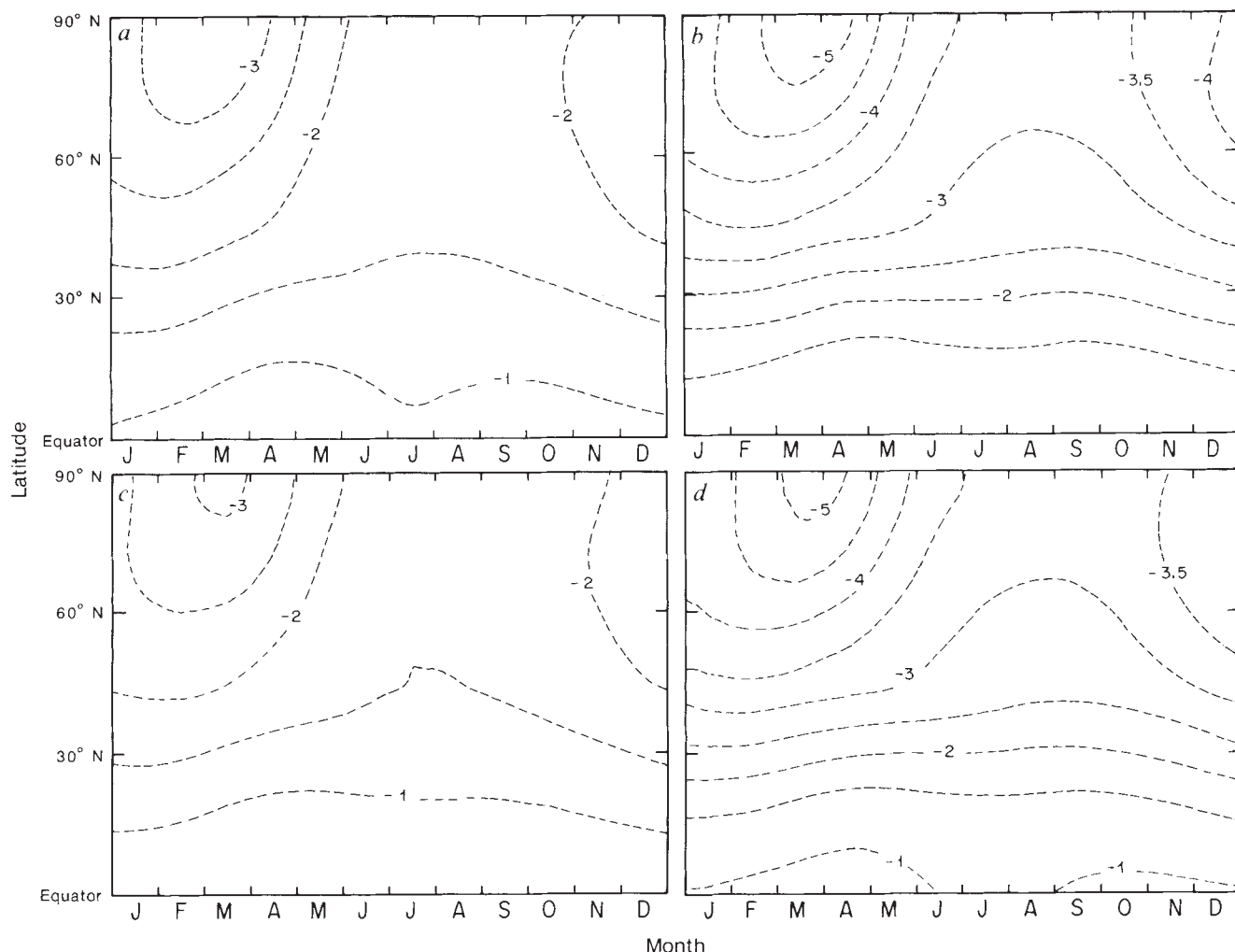


FIG. 2 Calculated decrease in column ozone between 1979 and 1990 (a and b) and 1990 and 2010 (c and d). In a and c we assume no heterogeneous

chemistry. Reaction (1) is incorporated in b and d, with reaction efficiency of 0.1.

efficiencies ranging from 0 to 0.1 are compared with different analyses of Dobson data in Table 1. Trends calculated at high winter latitudes with reaction (1) are 1.2–1.8 times as high as those calculated from models without heterogeneous chemistry. We note that somewhat larger enhancement in the calculated trends is obtained during the summer and at mid-latitudes.

Good agreement is obtained for winter trends from calculations including reaction (1) with efficiencies between 0.06 and 0.1. Recent analysis of 11 years (1979–90) of data from the total ozone mapping spectrometer (TOMS)² gives winter trends of ~ -6 to -8% per decade at northern mid-latitudes. These trends, taken at face value, are larger than both the Dobson results and our calculations. Trend analysis over this 11-year period also has errors of at least 50%. Analysis of satellite data over a longer period should reduce these uncertainties and determine the degree of agreement between calculations and observations.

Our calculated summer trends are higher than those derived from most of the analyses in Table 1, with the exception of trends derived from Dobson data between 1978 and 1988. This period was chosen in the analysis to correspond with the period of operation of the TOMS and SAGE instruments, and is not considered to be long enough to derive meaningful trends. Both the Dobson analysis and the satellite data indicate, however, that global trends, particularly during summer, are very sensitive to the period of data analysed, and vary considerably along a latitude band²⁵. The analysis of TOMS data in ref. 2 does indicate statistically significant negative trends during summer of ~ -2 to -4% per decade.

The model results in Fig. 1 and Table 1 indicate that the effect of reaction (1) does not increase much when the reaction efficiency changes from 0.06 to 0.1, and the mechanism seems to saturate for fast heterogeneous rate. As the net rate for reaction (1) increases so that the time constant approaches 1 day, the $\text{N}_2\text{O}_5/\text{NO}_x$ ratio is roughly inversely proportional to the heterogeneous rate. As a result, the net conversion rate of NO_x to HNO_3 becomes independent of the reaction efficiency, or, equivalently, of the aerosol loading. This result has important implications for the impact of reaction (1) after a volcanic eruption, or if there is an increasing trend in aerosol loading. If we consider only reaction (1) and assume a 25-fold increase in aerosol collision frequency in the Northern Hemisphere (0° to 50°N) in the year following the eruption of El Chichon, we calculate at most a 1.2% reduction in column ozone at mid-latitudes. The larger reduction calculated by other authors^{9,10} must be due to heterogeneous reactions involving ClNO_2 and HCl .

Our results illustrate that reaction (1) occurring in the global sulphate layer could explain a large fraction of the trends deduced from the Dobson and satellite data for the Northern Hemisphere, both at high- and mid-latitudes. Other mechanisms, which may or may not be related to the chlorine content, could also contribute to the observed ozone trends. Measurements of stratospheric constituents, particularly ClO , NO_x , NO_y , OH , HNO_3 and long-lived tracers, at different latitudes and seasons would further constrain the present model parameterizations of both transport and chemistry. Analysis of these data should include the details of the temperature and dependence on aerosol composition of this and other heterogeneous reactions for stratospheric conditions, and consider the effect of the motion of the sampled air masses on the photochemical behaviour of the trace species. $\square$

Received 7 January; accepted 28 May 1991.

1. WMO/UNEP Global Ozone Research and Monitoring Project Rep. No. 18, Ch. 7 (World Meteorological Organization, Geneva, 1990).
2. Stolarski, R. S., Bloomfield, P., McPeters, R. D. & Herman, J. R. *Geophys. Res. Lett.* **18**, 1015–1018 (1991).
3. Mozurkewich, M. & Calvert, J. G. *J. geophys. Res.* **93**, 15889–15896 (1988).
4. Van Doren, J. M. et al. *J. phys. Chem.* **95**, 1684–1689 (1991).
5. WMO/UNEP Global Ozone Research and Monitoring Project Rep. No. 20, Ch. 3 (World Meteorological Organization, Geneva, 1990).
6. Chipperfield, M. P. & Pyle, J. A. *J. geophys. Res.* **95**, 11865–11874 (1990).

7. Isaksen, I. S. A., Rognrud, B., Stordal, F., Coffey, M. T. & Mankin, W. G. *Geophys. Res. Lett.* **17**, 557–560 (1990).
8. Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Geophys. Res. Lett.* **15**, 257–260 (1988).
9. Hoffman, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5041 (1989).
10. Brasseur, G. P., Granier, C. & Walters, S. *Nature* **348**, 626–628 (1990).
11. Tolbert, M. A., Rossi, M. J. & Golden, D. M. *Geophys. Res. Lett.* **15**, 847–850 (1988).
12. Watson, L. R. et al. *J. geophys. Res.* **95**, 5631–5638 (1990).
13. Mather, J. H. & Brune, W. H. *Geophys. Res. Lett.* **17**, 1283–1286 (1990).
14. Turco, R. P., Toon, O. B., Hamill, P. & Whitten, R. C. *J. geophys. Res.* **86**, 1113–1128 (1981).
15. Hoffman, D. J. *Science* **248**, 996–1000 (1990).
16. Prather, M. J. & Watson, R. T. *Nature* **344**, 729–734 (1990).
17. Prather, M. J., McElroy, M. B., Wofsy, S. C. & Logan, J. A. *Geophys. Res. Lett.* **6**, 163–164 (1979).
18. Wofsy, S. C. *J. geophys. Res.* **83**, 364–378 (1978).
19. Ko, M. K. W., Sze, N. D., Livshits, M., McElroy, M. B. & Pyle, J. A. *J. Atmos. Sci.* **41**, 2381–2408 (1984).
20. Austin, J., Garcia, R. R., Russell, J. M. III, Solomon, S. & Tuck, A. F. *J. geophys. Res.* **91**, 5477–5485 (1986).
21. Brune et al. *Science* **242**, 558–562 (1988).
22. Tooney et al. *Geophys. Res. Lett.* **18**, 21–24 (1991).
23. Brune, W. H., Weinstock, E. M. & Anderson, J. G. *Geophys. Res. Lett.* **15**, 144–147 (1988).
24. Russell, J. M. III et al. *J. geophys. Res.* **93**, 1718–1736 (1988).
25. WMO/UNEP Global Ozone Research and Monitoring Project Rep. No. 20, Ch. 2 (World Meteorological Organization, Geneva, 1990).
26. Bojkov, R., Bishop, L., Hill, W. L., Reinsel, G. C. & Tiao, G. C. *J. geophys. Res.* **95**, 9785–9807 (1990).

ACKNOWLEDGEMENTS. We thank M. McFarland, S. R. Orfeo, M. J. Prather and S. R. Kawa for helpful discussions. Work was supported by the Upper Atmosphere Theory and Data Analysis program of NASA, the Fluorocarbon Program Panel of the Chemical Manufacturers Association, and Alternative Fluorocarbon Environmental Acceptability Studies (AFEAS).

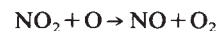
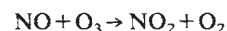
Importance of energetic solar protons in ozone depletion

Judy A. E. Stephenson & Malcolm W. J. Scourfield

Space Physics Research Institute, University of Natal, Durban 4001, South Africa

CHLORINE-catalysed depletion of the stratospheric ozone layer has commanded considerable attention since 1985, when Farman et al.¹ observed a decrease of 50% in the total column ozone over Antarctica in the austral spring. Here we examine the depletion of stratospheric ozone caused by the reaction of ozone with nitric oxide generated by energetic solar protons, associated with solar flares. During large solar flares in March 1989, satellite observations indicated that total column ozone was depleted by $\sim 9\%$ over $\sim 20\%$ of the total area between the South Pole and latitude 70°S . Chlorine-catalysed ozone depletion takes place over a much larger area, but our results indicate that the influence of solar protons on atmospheric ozone concentrations should not be ignored.

Although terrestrial sources exist for nitric oxide in the stratosphere², it can be generated by both solar protons and galactic cosmic rays^{3–5}. Solar protons of energies $>10\text{ MeV}$ enter the stratosphere at high geomagnetic latitudes ($\geq 60^\circ$) and can penetrate to altitude ranges where ozone is most abundant. The protons produce secondary electrons of energies $\sim 10\text{ eV}$, which ionize and dissociate molecular nitrogen leading, in turn, to the formation of nitric oxide⁶. The nitric oxide can then play an important part in ozone depletion through the following catalytic reactions:



Several different reactions can produce nitric oxide, and these have been used to calculate the nitric oxide formed during 'solar proton events' (SPEs) accompanying solar flares⁶. By definition, a solar proton event is said to have occurred when the flux of protons of energy $>10\text{ MeV}$ exceeds $10\text{ particles cm}^{-2}\text{ s}^{-1}\text{ sr}^{-1}$. More than 10 SPEs can occur in a year when the sun is very active, or there might be none at all in a quiet year⁷. Near solar maximum the proton fluxes can, on occasion, reach thousands of particles $\text{cm}^{-2}\text{ s}^{-1}\text{ sr}^{-1}$.

Large-scale stratospheric ozone depletion caused by solar protons has been observed⁸ using data from the backscattered ultraviolet experiment on the Nimbus 4 satellite (1970–77). Because of high background current on the diffuser plate,

however, ozone data were only available between 80° N and 80° S. Thus a large and important area of the polar cap was not covered by this data set. For the observations reported here, ozone data were obtained from the total-ozone-mapping spectrometer (TOMS) aboard the Nimbus 7 satellite launched in 1978. This experiment measures the global distribution of total column ozone daily. Each data point represents a mean taken

over an average cell size 60 km × 60 km as determined by the field of view of TOMS. Our data set was chosen from March 1989 during two SPEs involving large fluxes of high-energy protons (>10 MeV) of up to 3,500 and 2,000 particles $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$, respectively. Equally important, the solar illumination of the polar caps facilitated ozone measurements for most of the data set. The solar activity in March 1989 produced a great magnetic storm associated with large solar flares (some effects were the increased drag on satellites causing extensive orbit perturbation, failure of telecommunication systems and aurorae seen at unusually low latitudes of 30°).

To investigate any possible association between total column ozone and solar protons, the satellite ozone data were processed from March 5, the day before the two SPEs, through to March 31. In the results that follow, the total column ozone for March 5 is used as a baseline from which to gauge any effects from the SPEs beginning on March 8. Figure 1 shows colour plots of total column ozone, over an area extending from the south pole to 60° S, for March 5, 14 and 21. Each hue of the colour bar represents a range of 20 Dobson units (where 1 DU = 10^{-5} m of ozone at standard temperature and pressure), the lowest range being 180 to 199 DU. Attention is drawn to the region down to latitude 70° S, in which there is a marked increase after 5 March of the area covered by the deep blue colour corresponding to decreasing total column ozone. The other days between 5 and 21 March were also analysed and are consistent with a gradual increase in the area of ozone depletion.

This general decrease in total column ozone can be shown quantitatively by calculating the area covered by low total column ozone and expressing this as a percentage of the total area down to 70° S. These percentages are shown in Fig. 2 for low total column ozone in the ranges <200 DU, 200–219 DU and 220–239 DU. The increase in area covered by low total ozone is sustained until 26 March. We have also calculated the total mass of ozone (at STP), from the pole down to 70° S, taking into account the cosine dependence of the cell size of TOMS with latitude. The total mass decreased from 8.85×10^{10} kg on 5 March to 8.11×10^{10} kg on 21 March, representing a decrease of 7.4×10^9 kg or about 9%. An average total mass of ozone down to 70° S has also been calculated for the period 1 to 21 March in 1987 and 1988. These averages with standard deviations are, respectively, $9.26 \pm 0.24 \times 10^{10}$ kg and $9.12 \pm 0.29 \times 10^{10}$ kg. The decrease of 7.4×10^9 kg is therefore 2 to 3 times the standard deviation during these quiet times, and the observed decrease cannot be ascribed to natural variability.

Observations made from Sanae, Antarctica (70° 18' S, 2° 21' W) show that during March 1989 the temperatures at the

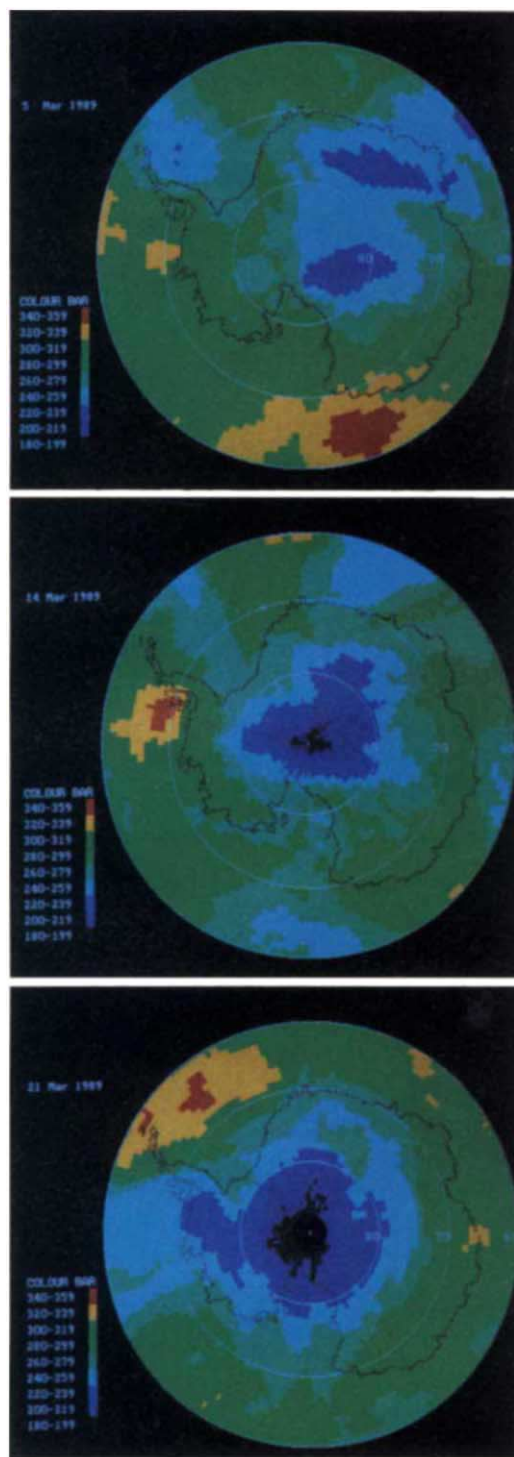


FIG. 1 Colour plots of total column ozone from the South Pole (90° S) to 60° S. Upper panel, before solar activity (5 March); middle and lower panels, during SPEs (14 March and 21 March respectively). The colour bar shows low total column ozone as deep blue, increasing through green and yellow to red. The overall range is 180 DU to 359 DU, and each hue represents a range of 20 DU. Latitudes 80° S and 70° S are shown, together with an outline of Antarctica.

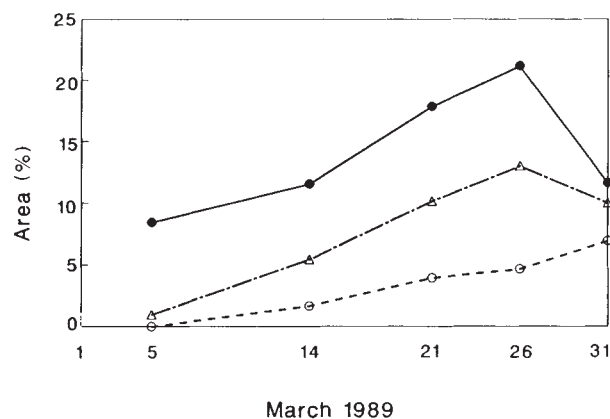


FIG. 2 The percentage area from the South Pole (90° S) to 70° S occupied by each of the three lower ranges of total column ozone (dark blues in Fig. 1), during selected days between 5 March and 31 March. Open circles, <200 DU; open triangles, 200–219 DU; full circles, 220–239 DU.

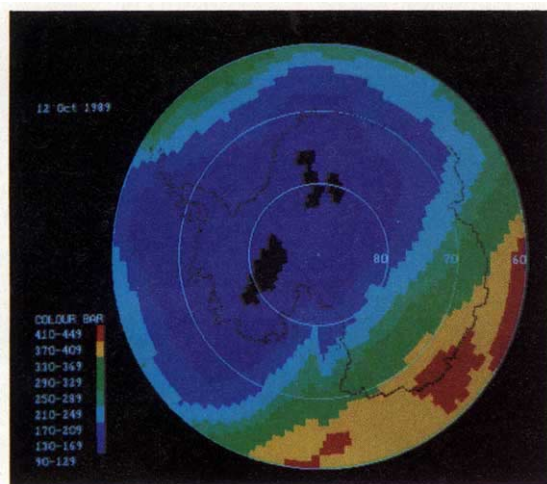


FIG. 3 The ozone hole on 12 October 1989. Because of an increased dynamic range of total column ozone the colour bar (90 DU to 410 DU) differs slightly from those of Fig. 1.

30 mbar level (an altitude of ~ 24 km) were, on average, about -40°C . On no occasion did the temperature fall to -80°C , which is the condition required for the formation of polar stratospheric clouds. Sanae was well within the region of low total column ozone on March 21, so the observed depletion cannot be due to the presence of such clouds. Heterogeneous reactions on the surface of ice particles in polar stratospheric clouds, with subsequent release of chlorine, can therefore be ruled out as a possible source of ozone depletion. This supports the idea that the depletion observed in March is due to solar protons.

We do not have any experimental observations from which to calculate the amounts of nitric oxide produced during these SPEs, but modelling has been done⁹ for a SPE in August 1972 which, coincidentally, had a similar particle profile to the March 1989 SPEs. The yield derived for nitric oxide, below 60 km and at geomagnetic latitudes greater than 60° , was 6×10^{15} molecules cm^{-2} . For comparison, Fig. 3 is a colour plot, again from the south pole to 60°S , for the springtime ozone hole on 12 October 1989. The colour bar is a little different from those of Fig. 1 because we have used an increased dynamic range of total column ozone values. By defining the region of ozone depletion as that area where the total column ozone is < 240 DU (the blue shades of Figs 1 and 3) we estimated the mass of ozone per unit area for 21 March and 12 October. These estimates are 4.6×10^{-3} kg m^{-2} for SPE-associated depletion and 3.6×10^{-3} kg m^{-2} for the chlorine-catalysed depletion. Chlorine-catalysed depletion, however, occurs over an area four times as great as that due to SPEs. Nevertheless, the influence of SPEs should not be ignored in any assessment of the Earth's ozone budget. $\square$

Fullerenes C_{60} and C_{70} in flames

Jack B. Howard, J. Thomas McKinnon,
Yakov Makarovskiy, Arthur L. Lafleur
& M. Elaine Johnson

Department of Chemical Engineering and Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA

THE fullerenes C_{60} and C_{70} were first identified¹ in carbon vapour produced by laser irradiation of graphite, and have recently been produced in macroscopic quantities²⁻⁵ by vaporization of graphite with resistive heating. It has also been suggested⁶⁻⁹ that fullerenes might be formed in sooting flames, and indeed all-carbon ions with mass/charge ratios suggestive of fullerenes have been detected in flames¹⁰⁻¹². These species were assumed to have the cage structures of fullerenes, but the mass spectroscopic evidence could not establish this conclusively. We have now collected samples of condensable compounds and soot from hydrocarbon combustion under a range of conditions, and analysed these using conventional techniques in an effort to detect fullerenes. Spectroscopic studies reveal the presence of C_{60} and C_{70} in yields and ratios that depend on temperature, pressure, carbon/oxygen ratio and residence time in the flame. Control of these conditions allows optimal yields of 3 g of fullerenes per kilogram of fuel carbon burned, and variation of the $\text{C}_{70}/\text{C}_{60}$ ratio over the range 0.26–5.7.

Premixed laminar flames of benzene and oxygen with argon diluent were stabilized on a water-cooled burner in a low-pressure chamber equipped with windows and feed-throughs for visual observation, optical diagnostics and probes, and exhausted into a vacuum pump. The flame is stabilized with a flat 70-mm-diameter front uniformly displaced from a drilled copper burner plate through which the feed mixture is delivered. The flame is surrounded by an annular nonsooting flame which provides a thermal shield, giving a nearly one-dimensional core within which temperature and species concentrations vary only with distance, or residence time, from the burner surface. The burner was previously used in mechanistic studies of soot nucleation and growth¹⁴, and the flames studied are of a type for which considerable data on temperature and chemical composition are available¹⁰⁻¹⁹.

Flames were produced under different sets of conditions over the following ranges: burner chamber pressure, 1.60–13.35 kPa; atomic carbon/oxygen ratio, 0.717–1.07; argon concentration, 0–39 mol %; gas velocity at the burner (298 K), 14–75 cm s^{-1} . Each flame was maintained for 53–170 min, during which a sample of condensable compounds and soot was withdrawn from the flame at a given distance from the burner using a quartz probe connected to a room-temperature filter, vacuum pump and gas meter. The mass of the sample was primarily that of soot, except at the lowest carbon/oxygen ratio where the flame was nonsooting. Soot was also collected from the inside surface of the burner chamber after each run. Using information on flame temperature and gas composition, the soot masses and gas volumes collected with the probe in the different sooting flames are found to correspond to soot yields in the range 0.2–12% of the carbon fed.

The samples were extracted with toluene, using an ultrasonic bath at room temperature, and filtered. The solution from one of the samples was evaporated to dryness and analysed with a mass spectrometer. The electron-impact mass spectrum is shown in Fig. 1. Comparing these data with those reported for fullerenes^{2-4,20}, we conclude that the soot sample contained a mixture of C_{60} and C_{70} fullerenes with molecular ions at mass/charge ratio, m/e , 720 and 840, respectively, and doubly charged molecular ions at m/e 360 and 420, respectively. This conclusion was confirmed by Fourier-transform infrared spectroscopy of the soot extract pressed in a KBr pellet, which

Received 20 December 1990; accepted 28 May 1991.

- Farman, J. C. *et al.* *World Meteorological Organization Rep.* No. 20, Vol. 1 (1990).
- Crutzen, P. J. *Q. J. R. Met. Soc.* **96**, 320–325 (1970).
- Warneck, P. *J. geophys. Res.* **77**, 6589–6591 (1972).
- Brasseur, G. & Nicolet, M. *Planet. Space Sci.* **21**, 939–961 (1973).
- Ruderman, M. A. & Chamberlain, J. W. *Planet. Space Sci.* **23**, 247–268 (1975).
- Crutzen, P. J., Isaksen, I. S. A. & Reid, G. C. *Science* **189**, 457–458 (1975).
- Hargreaves, J. K. *The Upper Atmosphere and Solar-Terrestrial Relations*, 232–243 (Van Nostrand Reinhold, Wokingham, 1979).
- Heath, D. F., Kreuger, A. J. & Crutzen, P. J. *Science* **197**, 886–889 (1977).
- Reagan, J. B. *et al.* *J. geophys. Res.* **86**, 1473–1494 (1981).

ACKNOWLEDGEMENTS. We thank R. P. McPeters and A. Krueger (NASA/GFSC), members of the TOMS Nimbus Experiment and Ozone Processing teams, and the National Space Science Data Center-A for Rockets and Satellites. Data on solar proton events were supplied by NOAA Space Environment Services Center.

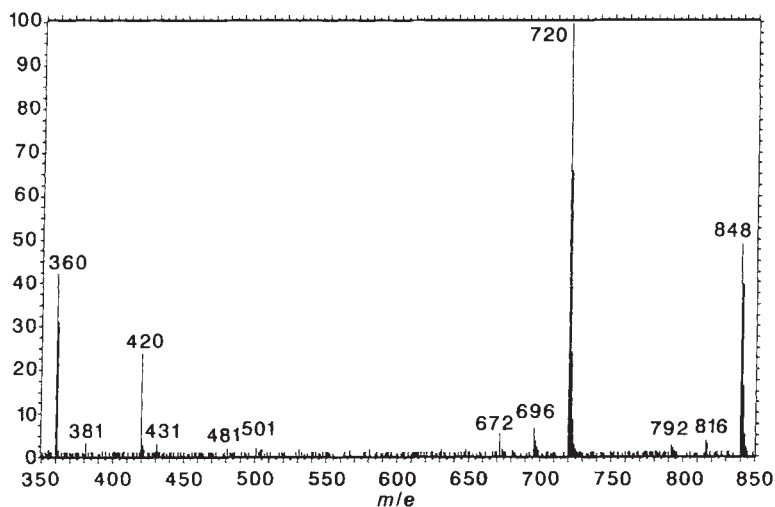


FIG. 1 Electron-impact mass spectrum of a flame soot extract. We used a direct-injection probe heated from 373 K to 673 K. Units are relative abundance (%).

gave strong absorption peaks consistent with those reported for fullerenes C_{60} (refs 2-5) and C_{70} (ref. 4).

The toluene extracts were fractionated with a high-performance liquid chromatograph (HPLC) coupled to a spectrophotometric diode-array detector. A separation scheme which has been proved effective for large polycyclic aromatic hydrocarbons (PAH)²¹ was used. A typical HPLC chromatogram for the toluene extracts is shown in Fig. 2. The broadband ultraviolet absorption shown is roughly proportional to mass for PAH²². The peaks labelled C_{60} and C_{70} gave ultraviolet spectra closely matching those published for C_{60} and C_{70} fullerenes, respectively^{2,4}. The peaks labelled A, B, C and D gave ultraviolet spectra that could not be traced to any known PAH, but seemed to be structurally related to the fullerenes. One or more of these satellite peaks are commonly present in chromatograms of soot extracts containing fullerenes.

To obtain broadband ultraviolet-visible spectra, solutions from HPLC fractionation of the soot extracts were concentrated by evaporation under nitrogen and the HPLC mobile phase was replaced with spectrograde decalin. The spectra of the ' C_{60} ' and the ' C_{70} ' peaks were acquired using a spectrophotometer. The results, illustrated in Fig. 3a and b, are virtually identical to those reported by Ajie *et al.*⁴.

Mass spectra of the HPLC fractions thought to be C_{60} and C_{70} fullerenes were acquired using the equipment and technique mentioned above. The ' C_{60} ' peak gave a mass spectrum with all the reported features of C_{60} fullerene^{2-4,20}. It gave a molecular-ion base peak at $m/e = 720$, showed no loss of hydrogen and had a significant peak for the doubly charged molecular ion at

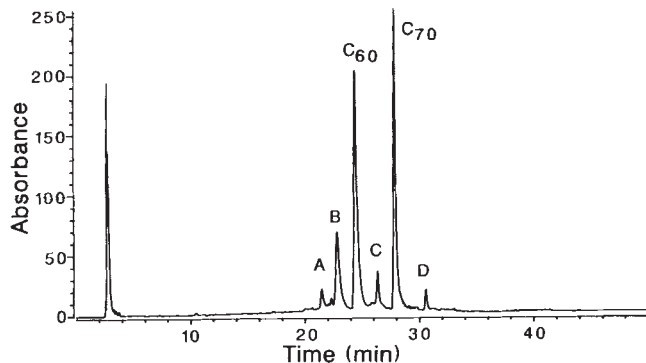


FIG. 2 High-performance liquid chromatogram of a flame soot extract. We used an octadecylsilyl-bonded silica column with a methanol-dichloromethane mobile phase. The absorbance scale gives the sum of absorbance over the 236-500-nm wavelength interval. Absorbance is given in milli-absorbance units.

$m/e = 360$. Similarly, the ' C_{70} ' peak gave a mass spectrum with features closely matching those of published spectra for C_{70} fullerene^{2-4,20}: a molecular-ion base peak at $m/e = 840$ and a doubly charged molecular ion at $m/e = 420$. Therefore, the identities of the HPLC peaks suggested by the ultraviolet-visible spectra were confirmed by the mass spectra. The HPLC method, including gravimetric calibration of the C_{60} and C_{70} peaks, was then used to analyse the toluene extracts of all the soot samples.

The mass of C_{60} plus C_{70} produced under the different sooting flame conditions is in the range 0.003%-9% of the soot mass, compared with 1-14% from graphite vaporization^{3-5,23}. The $(C_{60} + C_{70})$ yield, expressed as percentage of fuel carbon, varies from $2 \times 10^{-4}\%$ for a nonsooting flame to 0.3%, corresponding to 2.8 g of $(C_{60} + C_{70})$ per kg benzene burned, obtained at a pressure of 20 torr, a carbon/oxygen ratio of 0.995 with 10% argon, and a gas velocity at the burner of 49.1 cm s^{-1} (298 K). The flame temperature was $\sim 1,800 \text{ K}$. Given these results and the ability to scale up combustion reactors, flame synthesis is an alternative method for production of fullerenes.

The C_{70}/C_{60} molar ratio for the different flame conditions is in the range 0.26-5.7, compared with 0.02-0.18 for graphite vaporization¹⁻⁵. The ratio was 0.86 for the above conditions of maximum $(C_{60} + C_{70})$ yield. The much larger yields of C_{70} and the ability to control the C_{70}/C_{60} ratio by setting the flame conditions are important differences from the graphite vaporization technique. The C_{70}/C_{60} ratio tends to increase with increasing pressure, but other simple trends in this ratio are difficult to discern, perhaps because the overall change in the C_{70}/C_{60} ratio is a factor of 50 smaller than the overall change in the C_{60} and C_{70} yields, as well as couplings of the effects of the different independent variables.

The largest yields of fullerenes do not occur in the most heavily sooting flames. Also, the fullerene yield increases with increasing temperature or decreasing pressure under conditions where the same changes result in lower soot yields. These trends in the data reflect substantial differences between the formation mechanisms of fullerenes and those of soot. Both formation processes involve many of the same generic features, such as growth in small steps by addition of species contributing one to four carbon atoms and in large steps by reactions between polycyclic compounds, competition between growth and oxidative and unimolecular destruction reactions, and extensive involvement of radical sites at edge carbons in the growing polycyclic structures²⁴. The different trends shown by the two processes may arise because the various generic reactions are involved to different extents to form the very different structures. Soot formation does not involve the development of the curved or bowl-shaped structures required for fullerenes, nor would these curved structures permit the rapid stacking required in the

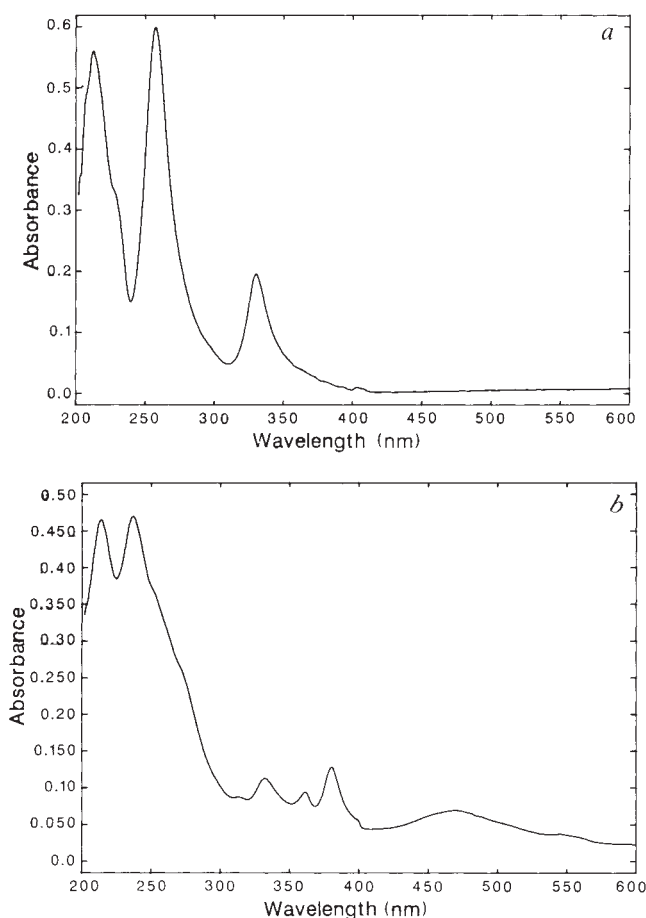


FIG. 3 *a*, Ultraviolet-visible spectrum of C_{60} fullerene peak. *b*, Ultraviolet-visible spectrum of C_{70} fullerene peak.

reactive coagulation²⁵ of the relatively flat sheet-like soot precursors²⁴. The observations that the very special fullerene structures are formed in flames, and that the formation occurs in the presence or absence of soot formation, provide new insight into the generic reactions important to both processes. □

ACKNOWLEDGEMENTS. We thank C. Zeng and C. Costello of the MIT Mass Spectrometry Facility for assistance in obtaining the mass spectra. The combustion research was supported by the Division of Chemical Sciences, Office of Basic Energy Research, US Department of Energy. The analysis was supported by the National Institute of Environmental Health Sciences. M.E.J. thanks the MIT Undergraduate Research Opportunities Program for a Dow Fellowship Award.

Change of coordination from tetrahedral gold-ammonium to square-pyramidal gold-arsonium cations

E. Zeller, H. Beruda, A. Kolb, P. Bissinger, J. Riede & H. Schmidbaur

Anorganisch-chemisches Institut der Technischen Universität München, Lichtenbergstrasse 4, D-8046 Garching, Germany

RECENT work¹⁻⁵ has shown that gold-based ligands in compounds of carbon and nitrogen can induce novel molecular structures with coordination numbers at C and N as high as 5 and 6. These phenomena must be ascribed to metal-metal interactions ($Au \cdots Au$), which can overrule bonding in classical configurations. Here we describe a study of the molecular structures of tetra(auro)ammonium ($(LAu)_4N^+$) and tetra(auro)arsonium ($(LAu)_4As^+$) cations (where L is a ligand). We find that the classical tetrahedral structure in these four-coordinate compounds is abandoned in favour of a square-pyramidal geometry once the radius of the central element is too large to allow for metal-metal bonding in a tetrahedral geometry. Thus, whereas the nitrogen compounds adopt a tetrahedral structure, for the larger arsenic atom an arsenic-capped square of gold atoms represents a more favourable core geometry. We have not yet been able to prepare the intermediate phosphorus compound, but we expect it also to have the square-pyramidal structure.

Methane, CH_4 , has a tetrahedral molecular structure, as first proposed by van't Hoff and LeBel in the final quarter of the nineteenth century. This prediction has been confirmed not only for this simplest hydrocarbon and its homologues (in fact for almost all organic molecules with tetracoordinate carbon atoms) but also for the isoelectronic ammonium cation NH_4^+ and all quaternary ammonium salts $NR_4^+X^-$ (where R is an organic or other substituent or ligand, and X is any counterion). The same geometrical concept is also valid for the silane and germane analogues (SiR_4 , GeR_4 and so on), and for quaternary phosphonium (PR_4^+) and arsonium cations (AsR_4^+). Lewis' simple 'octet rule', classical molecular orbital theory or modern 'ab initio' methods can be employed to rationalize this structural principle, which is perhaps best known as the model of 'sp³ hybridization' (see, for example, ref. 6).

Gold(I) complexes of the general form LAu^+ , where L is any neutral donor ligand, are known to have symmetry-related valence orbital characteristics similar to those of the proton H^+ or a carbo-cation R^+ , and in fact this 'isolobal' relationship is often borne out by close parallels in stoichiometry and structure of the corresponding derivatives⁷⁻⁹. It is therefore no surprise that the cations in the ammonium salts $[Au]_4N^+X^-$ have structures¹⁰ in which the nitrogen atom of the unit $(LAu)_4N^+$ is surrounded by four gold atoms in the form of a tetrahedron [$L=P(C_6H_5)_3$]. Although this tetrahedron was shown to be significantly distorted in earlier studies¹⁰⁻¹³ with the counterions $X=BF_4$ or PF_6 (termed structures **1a** and **1b**) our own investigations with the fluoride ($X=F$, **1c**) provided new data¹⁴ with much smaller deviations from the ideal geometry (Box 1 and Fig. 1; the compound was prepared from tris(trimethylsilyl)amine (or hexamethyldisilazane), triphenylphosphinegold chloride and caesium fluoride). For the cation with local crystallographic C_{3v} symmetry, the six Au-N-Au angles are split

Received 9 May; accepted 18 June 1991.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162-164 (1985).
- Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354-358 (1990).
- Taylor, R., Hare, J. P., Abdul-Sada, A. K. & Kroto, H. W. *JCS Chem. Commun.* 1423-1425 (1990).
- Ajje, H. et al. *J. phys. Chem.* **94**, 8630-8633 (1990).
- Haufler, R. E. et al. *J. phys. Chem.* **94**, 8634-8636 (1990).
- Zhang, Q. L. et al. *J. phys. Chem.* **90**, 525-528 (1986).
- Kroto, H. W. & McKay, K. *Nature* **331**, 328-331 (1988).
- Curl, R. F. & Smalley, R. E. *Science* **242**, 1017-1022 (1988).
- Kroto, H. *Science* **242**, 1139-1145 (1988).
- Gerhardt, P., Löffler, S. & Homann, K. H. *Chem. Phys. Lett.* **137**, 306-310 (1987).
- Gerhardt, P., Löffler, S. & Homann, K. H. *22nd Int. Symp. Combustion* 395-401 (The Combustion Institute, Pittsburgh, 1989).
- Löffler, S. & Homann, K. H. *23rd Int. Symp. Combustion* 355-362 (The Combustion Institute, Pittsburgh, 1990).
- Bittner, J. D. & Howard, J. B. *18th Int. Symp. Combustion* 1105-1116 (The Combustion Institute, Pittsburgh, 1981).
- McKinnon, J. T. thesis, Massachusetts Institute of Technology (1989).
- Homann, K. H., Mochizuki, M. & Wagner, H. G. Z. *phys. Chem.* **37**, 299-313 (1963).
- Homann, K. H., Morgeneyer, W. & Wagner, H. G. *Proc. Combustion Inst. Eur. Symp.* (ed. Weinberg, F. J.) 394-399 (Academic, London, 1973).
- Olson, D. B. & Calcote, H. F. *18th Int. Symp. Combustion* 453-462 (The Combustion Institute, Pittsburgh, 1981).
- Bockhorn, H., Fetting, F. & Wenz, H. *Ber. Bunsenges. phys. Chem.* **87**, 1067-1073 (1983).
- Howard, J. B. & Bittner, J. D. *Soot in Combustion Systems and Its Toxic Properties* (eds Lahaye, J. & Prado, G.) 57-91 (Plenum, New York, 1983).
- Luffer, D. R. & Schram, K. H. *Rapid Commun. Mass Spectrom.* **4**, 552-556 (1990).
- Fetzer, J. C. & Biggs, W. R. *J. Chromatogr.* **322**, 275-286 (1985); **346**, 81-92 (1985); **295**, 161-169 (1984).
- Lafleur, A. L., Monchamp, P. A., Plummer, E. F. & Wornat, M. J. *Anal. Lett.* **20**, 1171-1192 (1987).
- Krätschmer, W., Fostiropoulos, K. & Huffman, D. R. *Chem. Phys. Lett.* **170**, 167-170 (1990).
- Howard, J. B. *23rd Int. Symp. Combustion* (The Combustion Institute, Pittsburgh, in the press).
- McKinnon, J. T. & Howard, J. B. *Combust. Sci. Technol.* **74**, 175-197 (1990).

BOX 1. Crystal structure data

1c: $C_{72}H_{60}Au_4FNP_4$, $M_r=1,870.40$, trigonal, space group $P\bar{3}c1$ (No. 165), $a=b=17.925(3)$, $c=28.207(4)$ Å, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$, $V=7,848.8$ Å³, $Z=4$, $\rho_{\text{calc}}=1.582$ g cm⁻³, $\mu(\text{Mo } K_\alpha)=75.5$ cm⁻¹, $F(000)=5,088$, $T=23^\circ\text{C}$, 3,977 unique reflections to $(\sin \theta/\lambda)_{\text{max}}=0.6147$ Å⁻¹, 2,492 'observed' with $F_0 \geq 4.0 \sigma(F_0)$. Empirical absorption correction (relative transmission: 0.96–1.00); R (R_w)=0.069 (0.064), $\Delta\rho_{\text{fin}}(\text{max/min})=+1.84/-1.38$ e Å⁻³.

2a': $C_{72}H_{60}AsAu_4BF_4P_4 \cdot 3CH_2Cl_2$, $M_r=2,253.56$, monoclinic, space group $C2/c$ (No. 15), $a=34.804(4)$, $b=22.255(3)$, $c=26.948(4)$ Å, $\alpha=\gamma=90^\circ$, $\beta=133.81(1)^\circ$, $V=15,062.7$ Å³, $Z=8$, $\rho_{\text{calc}}=1.987$ g cm⁻³, $\mu(\text{Mo } K_\alpha)=85.3$ cm⁻¹, $F(000)=8,543.4$, $T=-56^\circ\text{C}$, 11,773 unique reflections to $(\sin \theta/\lambda)_{\text{max}}=0.5634$ Å⁻¹, 8,782 'observed' with $F_0 \geq 4.0 \sigma(F_0)$. Empirical absorption correction (relative transmission 0.82–1.00); R (R_w)=0.052 (0.041), $\Delta\rho_{\text{fin}}=+2.020/-1.650$ e Å⁻³.

2a'': $C_{72}H_{60}AsAu_4BF_4P_4 \cdot 4CH_2Cl_2$, $M_r=2,338.50$, triclinic, space group $P\bar{1}$ (No. 2), $a=14.533(2)$, $b=16.415(2)$, $c=18.817(2)$ Å, $\alpha=66.58(1)^\circ$, $\beta=78.02(1)^\circ$, $\gamma=88.05(1)^\circ$, $V=4,026.1$ Å³, $Z=2$, $\rho_{\text{calc}}=1.928$ g cm⁻³, $\mu(\text{Mo } K_\alpha)=80.45$ cm⁻¹, $F(000)=2,219.8$, $T=-56^\circ\text{C}$, 14,099 unique reflections to $(\sin \theta/\lambda)_{\text{max}}=0.5938$ Å⁻¹, 11,490 'observed' with $F_0 \geq 4.0 \sigma(F_0)$. Empirical absorption correction (relative transmission 0.75–1.00); R (R_w)=0.045 (0.048), $\Delta\rho_{\text{fin}}=+1.37/-1.92$ e Å⁻³. Solution of all structures was by direct methods (SHELXS-86). Refinement (SHELX-76) included all non-hydrogen atoms for **2a'** (786 parameters) and **2a''** (818 parameters); phenyl groups were treated as rigid groups for **1c** (78 parameters). Further details of the crystal structure investigations may be obtained from the Fachinformationszentrum Karlsruhe, D-7514 Eggenstein-Leopoldshafen 2, (Germany), on quoting the depository number CSD-55383, the names of the authors and the journal citation.

in groups of $3 \times 105.6(9)^\circ$ and $3 \times 113.0(8)^\circ$, and the Au...Au contacts (the six edges of the tetrahedron of gold atoms) in groups of $3 \times 3.227(2)$ and $3 \times 3.345(2)$ Å. The Au-N distances are $3 \times 2.005(9)$ and $1 \times 2.05(3)$ Å. These data are reminiscent of those found^{15,16} in alkyl or aryl(triauro)ammonium salts $RN(AuL)_3^+$, as far as the $N(AuL)_3$ unit of C_3 symmetry is concerned. The AuL apex in **1c** is clearly unique, however, in its long distance from the basal Au_3 plane. The results suggest that the $(LAu)_4N^+$ tetrahedron is not an ideal solution of the structure and is subject to almost random distortion depending on the nature of the counterion X.

In keeping with this finding, and contrary to the fact that the NH_4^+ cation does not add a proton to give the NH_5^{2+} dication, it has recently been shown that the cations in **1a** can accommodate a fifth LAu^+ unit to form the $(LAu)_5N^{2+}$ dication¹. There

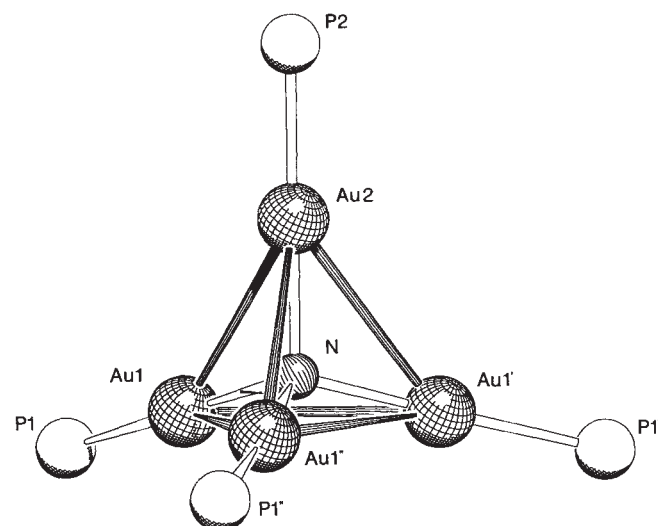


FIG. 1 Structure of the $N(AuP)_4$ core in crystals of the salt $[(C_6H_5)_3PAu]_4N^+ F^-$ **1c** with atomic numbering (phenyl groups and fluoride ions omitted).

is reason to ascribe the existence of this novel species with pentacoordinate (trigonal-bipyramidal) nitrogen to strong Au...Au interactions based on mixing of $Au(6s)$ and $(5d^{10})$ molecular orbital functions owing to relativistic effects. This concept soon led to the discovery¹⁷ of the phosphorus analogue $(LAu)_5P^{2+}$, although no classical $(LAu)_4P^+$ precursor has yet been isolated for this.

In the search for the corresponding arsenic compounds, the mixture of the reagents $[(CH_3)_3Si]_3As$ (ref. 18) and $[(C_6H_5)_3PAu]_4O^+ BF_4^-$ (ref. 19) in a molar ratio of 3:4, in tetrahydrofuran or dichloromethane as solvents, has now been found to give high yields of the expected arsonium salt $[(C_6H_5)_3PAu]_4As^+ BF_4^-$ (**2a**), with hexamethyldisiloxane and trimethylfluorosilane as the main byproducts. The corresponding chloride salt (**2b**) was obtained in a straightforward synthesis from $[(CH_3)_3Si]_3As$ and $(C_6H_5)_3PAuCl$ (molar ratio 1:4), with liberation of $(CH_3)_3SiCl$.

Compound **2a** was identified through microanalysis (values found were C 43.16%, H 3.04%, Au 39.55%; $C_{72}H_{60}AsAu_4BF_4P_4$ requires C 43.20%, H 3.03%, Au 39.42%) and fast-atom-bombardment mass spectrometry from 2-nitrobenzyl alcohol matrix: mass-to-charge ratio, $m/e = 1,911$ [$M(\text{cation})^+$, 100%]. NMR spectra of solutions in CD_2Cl_2 show a singlet resonance for ^{31}P ($\delta = 38.9$ p.p.m. relative to external 85% aqueous H_3PO_4) and only one set of phenyl 1H and ^{13}C signals, indicating equivalent $(C_6H_5)_3PAu$ groups. By contrast, the ^{197}Au Mössbauer spectrum (crystalline powder at 4 K) has an asymmetrical quadrupole doublet (isomeric shift 2.41, quadrupole splitting 7.11 mm s⁻¹), suggesting a slightly different environment of the gold atoms in the solid.

Compound **2a** crystallizes in no less than three different forms from mixtures of CH_2Cl_2 with pentane or diethylether, with different contents of solvent molecules and with colours ranging from pale yellow and orange to burgundy red. Redissolved in CH_2Cl_2 , these crystals give solutions indistinguishable from each other by NMR spectroscopy. Single-crystal X-ray diffraction analyses of two of these forms (**2a'** and **2a''**; see Box 1) show that both contain the same ionic components, $[(C_6H_5)_3PAu]_4As^+$ and BF_4^- , but with different numbers of CH_2Cl_2 molecules and in somewhat different packing. (The structure of the third form follows the same principle, but at present disorder problems prevent a detailed evaluation at the same level.)

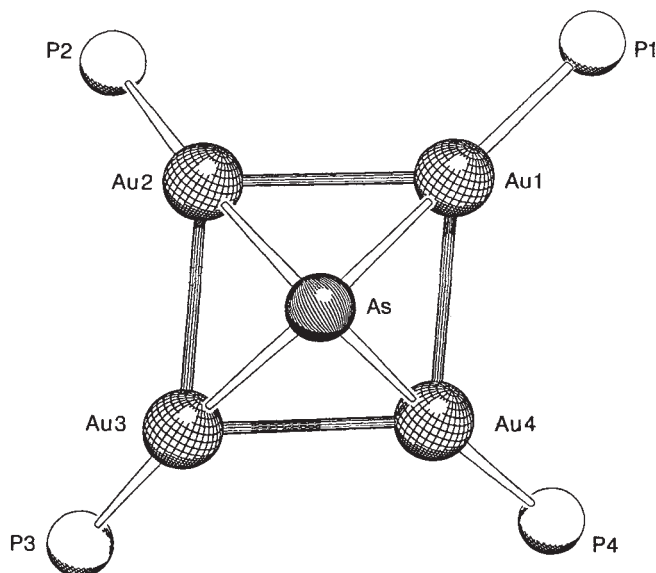


FIG. 2 Structure of the $As(AuP)_4$ core in crystals of the salt $[(C_6H_5)_3PAu]_4As^+ BF_4^- \cdot 4CH_2Cl_2$ (**2a''**) with atomic numbering (phenyl groups, dichloromethane molecules and tetrafluoroborate anions omitted).

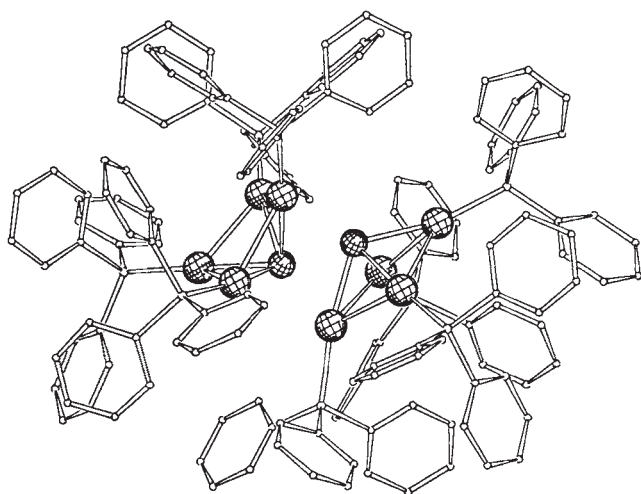


FIG. 3 Packing of two cations $[(C_6H_5)_3PAu]_4As^+$ in crystals of the tetrafluoroborate salt **2a''** (tetrafluoroborate anions and dichloromethane molecules omitted).

Against all expectations, the cations have a tetragonal pyramidal structure, with the arsenic atom occupying the apex above a slightly distorted square of gold atoms (Fig. 2). The coordination geometry at each gold atom (P–Au–As) is nearly linear, and the average apical Au–As–Au angle is 70.7° for both **2a'** and **2a''**, with an average As–Au distance of $2.50 \text{ \AA}$ ($2.49 \text{ \AA}$) and (probably the most important point) average intramolecular Au...Au contacts at $2.90 \text{ \AA}$ ($2.89 \text{ \AA}$). In the crystals, the cations appear in pairs (Fig. 3), with two longer intermolecular As...Au and Au...Au contacts in **2a'** (two Au...Au contacts in **2a''**) of 3.117 and $3.672 \text{ \AA}$ ($2 \times 3.414 \text{ \AA}$), respectively. These are probably responsible for some of the structural distortions of the individual units, for the splitting of the ^{197}Au Mössbauer lines and for the colour changes. There is ample evidence, however, that these contacts are not retained in solution (colour, NMR) or in the gas phase (fast-atom-bombardment mass spectrometry). The monomers are related by C_2 and C_i point group symmetry in **2a'** and **2a''**, respectively.

There is no precedent for such an unusual structure of molecules with a tetra-coordinate central atom and eight valence electrons as part of a flexible, nonforcing skeleton²⁰. If only radial As–Au bonding is considered, and a lone pair of electrons is allocated to the As apex, then the $AsAu_4$ core has to be described as electron-deficient with only three bonding molecular orbitals filled under local C_{4v} symmetry. It is very likely, however, that peripheral Au...Au bonding strongly contributes to the cluster stability, as indicated by the short Au...Au distances of less than $3.00 \text{ \AA}$ at the base of the As-capped pyramid. Taking the As–Au distances of $2.50 \text{ \AA}$ as a basis, a simple calculation shows that the Au...Au contacts in an As-centred tetrahedron would be $\sim 4.08 \text{ \AA}$, and thus far too long for efficient Au...Au interactions^{2,21}. In the nitrogen-centred tetrahedron (**1c**), on the other hand, the small atomic radius of nitrogen (as compared with arsenic) still allows Au...Au contacts of $\sim 3.25 \text{ \AA}$ and thus makes a rearrangement to a pyramidal geometry unnecessary.

It thus seems that the concept of bonding Au...Au interactions between seemingly closed-shell ($5d^{10}$) configurations, broken up through relativistic effects ('aurophilicity'²²), provides the solution of what otherwise could have been yet another paradox in the chemistry of metallo-complexes of gold. It has been pointed out previously^{2,22,23} that many of the special features in gold chemistry have their origin in relativistic effects, which are known to be a local maximum for this element. As the velocity of the electrons in the electric field of very highly charged nuclei (such as gold, $Z = 79$) approaches the velocity of light,

considerable changes in their relative energy levels arise, which reduce the gap particularly between the $6s/6p$ and $5d$ states, thus allowing additional bonding not foreseen in classical terms. More detailed theoretical work, already successful^{22–27} for a description of the bonding in hypercoordinate carbon atoms^{3–5} of the types $(LAu)_6C^{2+}$ and $(LAu)_5C^+$, is in progress. □

Received 27 February; accepted 9 May 1991.

- Grohmann, A., Riede, J. & Schmidbaur, H. *Nature* **345**, 140–142 (1990).
- Schmidbaur, H. *Gold Bull.* **23**, 11–20 (1990).
- Scherbaum, F., Grohmann, A., Huber, B., Krüger, C. & Schmidbaur, H. *Angew. Chem. Int. Edn Engl.* **27**, 1544–1546 (1988).
- Scherbaum, F., Grohmann, A., Müller, G. & Schmidbaur, H. *Angew. Chem. Int. Edn Engl.* **28**, 463–465 (1989).
- Scherbaum, F., Huber, B., Müller, G. & Schmidbaur, H. *Angew. Chem. Int. Edn Engl.* **27**, 1542–1544 (1988).
- Greenwood, N. N. & Earnshaw, A. N. *Chemistry of the Elements* (Pergamon, Oxford, 1986).
- Hoffmann, R. *Angew. Chem. Int. Edn Engl.* **21**, 711–723 (1982).
- Stone, F. G. A. *Angew. Chem. Int. Edn Engl.* **23**, 85–95 (1984).
- Hall, K. P. & Mingos, D. M. P. *Prog. inorg. Chem.* **32**, 237–254 (1984).
- Slovokhotov, Yu. L. & Struchkov, Yu. T. *J. organometall. Chem.* **277**, 143–146 (1984).
- Perevalova, E. G., Smyslova, E. I., Dyadchenko, V. P., Grandberg, K. I. & Nesmeyanov, A. N. *Izv. Akad. Nauk SSSR Ser. Khim.* **1455** (1980) (in Russian).
- Brodbeck, A. thesis, Univ. of Tübingen (1990).
- Brodbeck, A. & Strähle, J. *Acta Crystallogr. A* **46**, C-232 (1990).
- Kolb, A. & Bissinger, P. thesis, Techn. Univ. München (1990).
- Grohmann, A., Riede, J. & Schmidbaur, H. *J. chem. Soc. Dalton Trans.* 783–788 (1991).
- Ramamoorthy, V. & Sharp, P. R. *Inorg. Chem.* **29**, 3336–3340 (1990).
- Weidenhiller, G., Steigelmann, O., Riede, J. & Schmidbaur, H. *Angew. Chem. Int. Edn Engl.* **29**, 433–435 (1991).
- Becker, G., Gutekunst, G. & Wessely, H. J. *Z. anorg. allg. Chem.* **462**, 113–129 (1980).
- Nesmeyanov, A. N. et al. *J. organometall. Chem.* **201**, 343–349 (1980).
- Schmidbaur, H., Aly, A. A. M. & Schubert, U. *Angew. Chem. Int. Edn Engl.* **17**, 846–847 (1978).
- Jones, P. G. *Gold Bull.* **19**, 46–57 (1986).
- Rösch, N., Görling, A., Ellis, D. E. & Schmidbaur, H. *Angew. Chem. Int. Edn Engl.* **28**, 1357–1359 (1989).
- Rösch, N., Görling, A., Ellis, D. E. & Schmidbaur, H. *Inorg. Chem.* (in the press).
- Mingos, D. M. P. *J. Chem. Soc. Dalton Trans.* 1163–1170 (1976).
- Evans, D. G., Mingos, D. M. P. *J. organometall. Chem.* **232**, 171–174 (1982).
- Mingos, D. M. P. *Nature* **345**, 113–114 (1990).
- Mingos, D. M. P. & Kanter, R. P. F. *J. organometall. Chem.* **384**, 405–415 (1990).

ACKNOWLEDGEMENTS. We thank F. E. Wagner and F. H. Kreissl for recording the Mössbauer and mass spectra, respectively. This work was supported by the Deutsche Forschungsgemeinschaft (Leibniz-Programm), the Fonds der Chemischen Industrie, and by Degussa AG and Hereaus GmbH.

Forces between polymer-bearing surfaces undergoing shear

Jacob Klein, Dvora Perahia & Sharon Warburg

Polymer Research Department, Weizmann Institute of Science, Rehovot 76100, Israel

ADSORBED or grafted polymers are used to control phenomena such as colloidal stability, fluid flow and the tribological properties of surfaces. Forces between polymer-bearing surfaces have been studied comprehensively over the past decade^{1–10}, but little is known about such forces in shear. These are intimately related to the dynamic as well as the equilibrium properties of the surface-attached chains. We have constructed a device that measures directly the forces that act between surfaces bearing polymer layers in a liquid medium as they slide past each other. For the case of mica sheets bearing end-grafted chains of polystyrene in toluene (a good solvent), we find that, as the surfaces move parallel to each other, there is a marked change in the normal forces between them. These become increasingly repulsive at higher velocities. The effect occurs only above certain critical shear rates, probably related to relaxation dynamics of the end-grafted chains themselves. Our findings have direct implications for the properties of polymeric lubricants, and for the rheological behaviour both of stabilized dispersions and of multi-phase polymeric systems.

Our apparatus, shown schematically in Fig. 1, can measure simultaneously both the normal forces $F_{\perp}(D)$ and the lateral (shear) forces $F_{\parallel}(D)$ between curved mica surfaces as they slide past each other a closest distance D apart. It resembles previous approaches used to investigate frictional and viscous forces

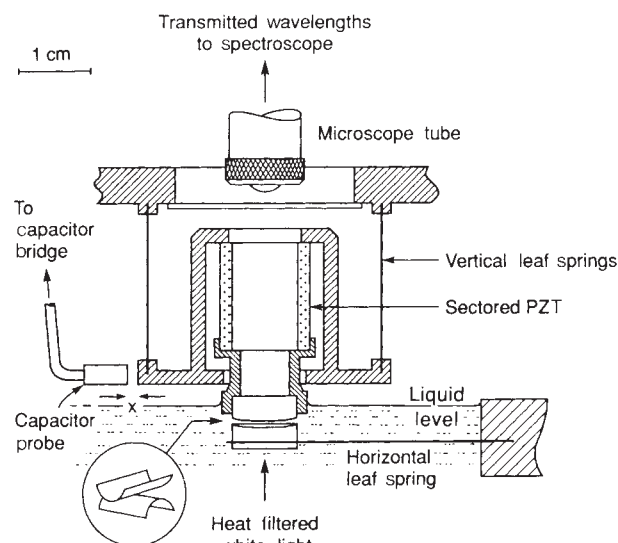


FIG. 1 Schematic diagram of measuring apparatus. The closest distance D between curved mica surfaces immersed in liquid is measured to ± 0.2 nm using optical interferometry^{1,17,18}, and is controlled by a three-stage mechanism. The fine stage is the sectorized piezoelectric tube (Vernitron PZT 8-8031-5H), mounted on vertical springs, which can move the top mica surface normally and laterally¹⁶ (to ± 0.1 nm in the x direction) relative to the lower surface. Lateral movements of up to $3 \mu\text{m}$ are possible, either monotonically, or at frequencies up to 540 Hz, or by means of a step pulse. Suitable coupling of the potential of the PZT inner surface to that of its outer sectors ensures parallel motion of the surfaces (less than 1 nm change in D over $2 \mu\text{m}$ lateral displacement). Bending of the vertical springs is monitored by an air-gap capacitor (MIA, amplifier AS-1021 PA/06, Probe ASP-1-LA) to within $\delta x = \pm 0.3$ nm. Stiffnesses K_1 and K_2 of the vertical and horizontal leaf-springs are $3.0 \times 10^2 \text{ N m}^{-1}$ and $1.5 \times 10^2 \text{ N m}^{-1}$ respectively, providing a resolution $K_1 \delta x$ of better than 10^{-7} N in measuring the lateral forces. The surfaces and lower spring are immersed in a small stainless steel bath (volume 16 ml, not shown) enclosed in a larger stainless steel box; all parts in contact with liquid are stainless steel, glass or chlorotrifluoroethylene.

between sliding mica sheets close to contact¹¹⁻¹⁵. The present device has been designed to measure the very different effects associated with the shear of macromolecular surface phases. A sectorized piezoelectric tube (PZT) is mounted on vertical leaf-springs and used to move the upper surface either laterally¹⁶ or normally relative to the lower one which is mounted on a horizontal leaf-spring. Lateral forces between the surfaces bend the vertical springs, and this bending is measured (to within ± 0.3 nm) using an air-gap capacitor. Normal forces between them are monitored through the bending of the horizontal spring, using optical interferometry as in earlier studies^{1-5,17,18}. Extremely parallel motion of the surfaces over a wide range of

velocities is possible using the sectorized PZT; in particular, this allows us to measure the coupling between shear and normal forces. Overall, we can measure normal and lateral forces with comparable resolution and sensitivity.

Normal forces $F_{\perp}(D)$ between the curved mica sheets without shearing were measured as described in earlier studies^{4,5}, both before and after coverage of the surfaces by polymeric chains from the surrounding toluene medium. The range and magnitude of these forces (Fig. 2a) are closely similar to values earlier reported, and reveal unambiguously the tethering of the polymer to the mica surface by the zwitterion at its end^{4,5}. Such profiles, which also indicate the amount of polymer on each surface,

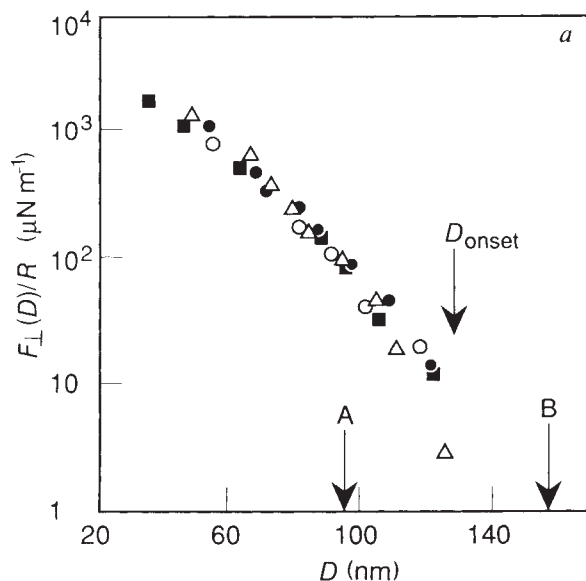
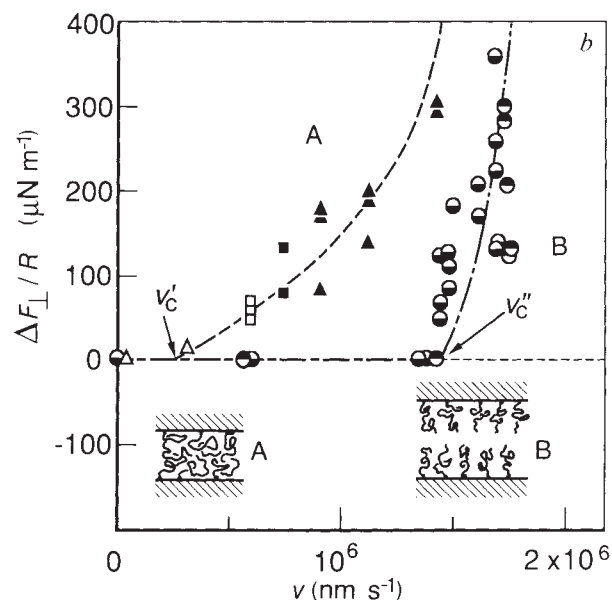


FIG. 2 a, Normal force profile $F_{\perp}(D)/R$ between curved mica sheets (mean radius of curvature $R \approx 1$ cm) a closest distance D apart. Sucrose was used to glue the sheets to the glass formers. The surfaces are immersed in a solution 0.01 wt% of zwitterion-terminated polystyrene, PS-X, in toluene (Fluka, spectroscopic grade) at 23 ± 1 °C and are covered by layers of the polymer, anchored to each surface by the zwitterion group ($\text{N}^+(\text{CH}_3)_2(\text{CH}_2)_3\text{SO}_3^-$). Different symbols are from measurements carried out before and after shear of the surfaces. $D_{\text{onset}} = 125 \pm 5$ nm indicates the surface separation for onset of monotonic repulsion (the separation at which the layers come into contact), and A and B indicate the separations at which the data of Fig. 2b were taken. The PS-X has² a weight-average molecular



weight 1.41×10^5 and polydispersity 1.02. b, Variation of the increment $\Delta F_{\perp}/R$ in the normal force between mica surfaces a fixed distance D apart, bearing terminally anchored PS-X layers, as a function of their mean lateral velocity v . Curves A and B are for $D = 94.5 \pm 1$ nm and $D = 155.2 \pm 1$ nm respectively, corresponding to the separations A and B indicated in Fig. 2a. The lateral motion is sinusoidal with frequency ν and amplitude Δx , and the velocity is evaluated as $v = 2\nu\Delta x$. In curve A, Δ is for $\nu = 200$ Hz; $\blacktriangle$, $\square$ for $\nu = 400$ Hz; $\blacksquare$ for $\nu = 500$ Hz. In curve B, $\bullet$ and $\circ$ are both at $\nu = 400$ Hz, from different runs. The continuous broken curves are a guide to the eye. The drawings (bottom) indicate schematically the polymer layers at the separations A and B when the surfaces are at rest ($v=0$).

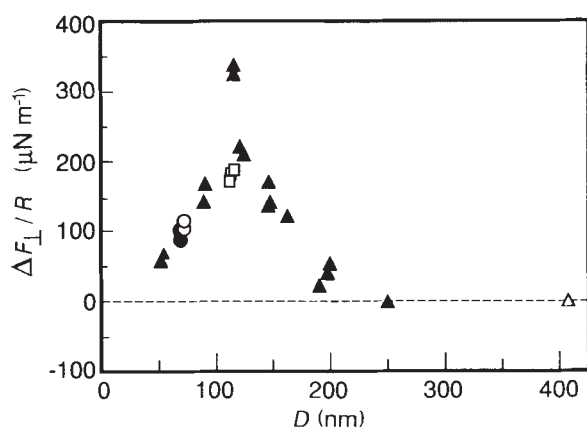


FIG. 3 Variation of the normal force increment $\Delta F_{\perp}/R$ at a fixed value of the mean shear velocity between mica surfaces bearing the terminally-anchored PS-X chains, as a function of their separation, D . Different symbols correspond to different runs. The motion is provided by a periodic step-pulse (half-rise time $t_{1/2} = 6 \times 10^{-5}$ s) of amplitude $\Delta x = 315$ nm, which may be considered to correspond to a mean velocity $v = \Delta x / 2t_{1/2} = 2.6 \times 10^6$ nm s $^{-1}$. Because of the different waveforms (step pulse and sinusoidal, respectively) used to generate the lateral motion, it is not appropriate to compare this mean shear velocity directly with those in Fig. 2b.

were frequently determined during an experiment to check the integrity of the grafted layers, especially after shear. They are identical within scatter to the pre-shear profile, and indicate that shear in these experiments does not remove polymer from the anchoring mica substrates.

We investigated the effect on the normal forces $F_{\perp}(D)$ of sliding the top mica surface relative to the bottom one as a function of D and of the relative sliding (or shear) velocity v . For the bare mica surfaces in toluene, no change in $F_{\perp}(D)$ was detected down to surface separations of ~ 5 nm (at which point the surfaces jumped into strong adhesive contact⁴) over the entire accessible range of shear frequencies and velocities (see Fig. 1). In contrast, following the anchoring of polystyrene chains on each mica surface, we observe considerable coupling between the lateral motion and the normal forces $F_{\perp}(D)$ between the polymer-bearing surfaces. The horizontal spring (stiffness K_2) supporting the lower mica surface bends by ΔD when the upper surface moves at sufficiently high velocity parallel to it. Such a bending reflects a change $\Delta F_{\perp} = K_2 \Delta D$ in the normal force acting between the surfaces.

The coupling is clearly seen in Fig. 2b. As v increases, little change is observed up to a certain critical velocity (marked v_c' , v_c'' for the two values of D shown in Fig. 2), beyond which the additional repulsion $\Delta F_{\perp}$ grows rapidly; this behaviour is reversible on changing v . The increase in repulsion beyond the critical shear velocities seems to be due to lateral stretching of the surface-tethered polymer chains when they shear sufficiently rapidly past each other. At low shear velocities and moderate compressions (as for the left-hand curve in Fig. 2b), little lateral force $F_{\parallel}(D)$ is measurable between the polymer-bearing surfaces (J. K. and D. P., manuscript in preparation); once the shear rates exceed the relaxation rate of a tethered chain or of a part of it, however, there is an increased tendency to stretch it laterally. Such stretching of chains in the compressed layers is expected to increase the extent to which monomers in the gap interact and repel each other^{19,20}. This in turn increases the local osmotic pressure²⁰ and is manifested as an increase in $\Delta F_{\perp}$.

Repulsion sets in suddenly at the surface separation $D = 155$ nm (right-hand curve in Fig. 2b). This separation is considerably greater than the separation $D_{\text{onset}} = 125$ nm for contact between the polymer layers (Fig. 2a), so that at rest, a substantial fluid gap separates them as indicated. The sudden increase in $\Delta F_{\perp}$ in this case (and in others, not shown, for which $D > D_{\text{onset}}$) suggests the following: when fluid flows rapidly enough ($v > v_c''$) past the polymer layers, it stretches the end-tethered chains, and the layers become thicker as a result (calculations by J.-L. Barrat, personal communication). This thickening narrows the gap between them and increases the effective shear rate of the liquid in the gap, which leads to further thickening of the two polymer layers, and so on. This cooperative effect rapidly results in contact between the opposing layers, and their mutual repulsion, because of monomeric interactions², causes the sharp rise in $\Delta F_{\perp}$.

The variation of $\Delta F_{\perp}$ with surface separation D at a fixed value of the mean relative velocity v is shown in Fig. 3. At large separations (≥ 200 nm) there is no measurable repulsion between the surfaces, but as D decreases, an increasing repulsive force $\Delta F_{\perp}$ is induced by the shearing motion, as described above. $\Delta F_{\perp}$ goes through a maximum (close to $D = D_{\text{onset}}$) before decaying as the layers are progressively compressed. This behaviour is reversible on changing D , and the maximum in $\Delta F_{\perp}$ is found also at other (sufficiently high) shear velocities. This maximum in $\Delta F_{\perp}$ as the surfaces approach is due to interplay of several factors but may be qualitatively understood as follows: at large separations the shear rate in the gap is low, and any stretching of chains is not sufficient to lead to thickening of and repulsion between the layers. At lower separations ($D \leq 200$ nm for the relative shear velocity in Fig. 3) such repulsion, manifested as $\Delta F_{\perp}$, does take place and increases as D decreases, as does the effective shear rate in the gap (as in Fig. 2b). At still higher compressions ($D \leq D_{\text{onset}}$), the increasing monomer concentration in the gap reduces the efficiency of the chain-stretching in promoting repulsive monomeric interactions²⁰. This leads eventually to the observed reduction in the repulsion $\Delta F_{\perp}$.

Our results show that the normal forces between surfaces bearing grafted polymer layers in a good solvent increase markedly when the compressed layers are sheared, and that an uncompressed layer increases its thickness when fluid flows past it. These effects occur above certain critical shear rates, probably related to relaxation rates of the end-tethered chains. They may provide a new approach for studying the dynamic behaviour of confined macromolecules. □

Received 4 April; accepted 24 May 1991.

- Klein, J. *Nature* **288**, 248–250 (1980).
- Klein, J. *Molecular Conformation and Dynamics of Macromolecules in Condensed Systems* (ed. Nagasawa, M.) 333–352 (Elsevier, Amsterdam, 1988).
- Patel, S. & Tirrell, M. *Ann. Rev. Phys. Chem.* **40**, 597–635 (1989).
- Taunton, H. J., Toprakcioglu, C., Fetters, L. J. & Klein, J. *Nature* **332**, 712–714 (1988).
- Taunton, H. J., Toprakcioglu, C., Fetters, L. J. & Klein, J. *Macromolecules* **23**, 580–589 (1990).
- de Gennes, P. G. *Macromolecules* **15**, 492–501 (1982).
- Klein, J. & Pincus, P. *Macromolecules* **15**, 1129–1137 (1982).
- Scheutjens, J. & Fleer, G. *Macromolecules* **18**, 1882–1894 (1985).
- Alexander, S. *J. Phys.* **38**, 983–989 (1977).
- de Gennes, P. G. *Adv. Colloid Interf. Sci.* **27**, 189–207 (1987).
- Bailey, A. I. *J. appl. Phys.* **32**, 1407–1413 (1961).
- Israelachvili, J. N. & Tabor, D. *Wear* **24**, 386–390 (1973).
- Briscoe, B. J. & Evans, D. C. B. *Proc. R. Soc. Lond. A* **380**, 389–398 (1982).
- van Alsten, J. & Granick, S. *Phys. Rev. Lett.* **61**, 2570–2573 (1988).
- Homola, A. M. Israelachvili, J. N., Gee, M. L. & McGuiggan, P. M. *J. Tribology* **111**, 675–682 (1989).
- Carr, R. G. *J. Microsc.* **152**, 379–388 (1988).
- Tabor, D. & Winterton, R. H. S. *Proc. R. Soc. Lond. A* **312**, 435–450 (1969).
- Israelachvili, J. N. & Adams, G. *J. Chem. Soc. Faraday I* **74**, 974–1001 (1978).
- Pincus, P. *Macromolecules* **9**, 386–392 (1976).
- Rabin, Y. & Alexander, S. *Europhys. Lett.* **13**, 49–54 (1990).

ACKNOWLEDGEMENTS. J. K. thanks C. Quate and E. J. Kramer for their suggestions. We thank L. J. Fetters for the PS-X polymer, and R. deRoos and Y. Gat for technical assistance. We appreciate discussions with S. Alexander, J.-L. Barrat, Y. Rabin, A. Silberberg and particularly F. Pincus. This work was supported by the Minerva Foundation, the National Council for Research and Development (Israel) and the Forschungszentrum Juelich (GMBH). J.H. is the incumbent of the Herman Mark chair of Polymer Physics.

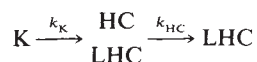
The stability of hydrocarbons under the time-temperature conditions of petroleum genesis

Frank D. Mango

Shell Development Company, Bellaire Research Center, PO Box 481, Houston, Texas 77001, USA

THE contention that hydrocarbons are unstable under the time-temperature conditions of petroleum generation (catagenesis) enjoys broad acceptance¹⁻⁶. At temperatures in the region of 100–150 °C, hydrocarbons are believed to decompose progressively over geological time to lighter hydrocarbons, ultimately methane and pyrobitumen^{7,8}. There are geological contradictions to this view⁹, however, and the recent finding¹⁰ that the cycloalkane ring should remain stable for billions of years under catagenic conditions demands a review of its underlying assumptions. For example, are ordinary hydrocarbons unstable under catagenic conditions? Are the light hydrocarbons, including methane, produced through the thermal decomposition of higher-molecular-weight hydrocarbons? Here I address these questions from two independent perspectives. First, the relative stabilities of hydrocarbons and their kerogenous precursors are studied experimentally. Second, the compositions of natural petroleum deposits are analysed for evidence of thermal decomposition. The results suggest that the hydrocarbons in petroleum should be at least three orders of magnitude more stable than their kerogenous precursors under catagenic conditions, and that natural deposits of petroleum and gas do not contain cycloalkane and isoalkane contents indicative of progressive thermal decomposition to gas. Thus the assumption that hydrocarbons are unstable under catagenic conditions and progressively decompose to methane and pyrobitumen seems to be incorrect.

In the general view of petroleum generation⁶, the conversion of kerogen (K) into liquid hydrocarbons (oil) and gas is a progressive process, with K thermally decomposing to hydrocarbons (HC) and light hydrocarbons (LHC) in the primary step, and HC undergoing further decomposition to LHC in the secondary step:



If we assume that the kerogenous precursors to hydrocarbons are higher-molecular-weight hydrocarbons, then the rate constants for kerogen and hydrocarbon decompositions (k_K and k_{HC} , respectively) should be similar under catagenic conditions. For example, normal alkanes and linear polyethylene should decompose at similar rates, as the bond energies in the key bond-breaking steps are essentially the same in both cases, assuming other kinetic factors remain equal. Consistent with this, the first-order rate constants for the decompositions of linear polyethylene¹¹ and normal heptane¹² are roughly equal at temperatures around 150 °C. A general parity of rates is implicit in the assumption that hydrocarbons undergo thermal decomposition to lighter hydrocarbons under the time-temperature conditions of petroleum generation. Here I analyse this assumption through kinetic studies on the thermal decomposition of hydrocarbons and their presumed kerogenous precursors under identical experimental conditions.

The asphaltene fraction (As) of bitumen is regarded by many as a low-molecular-weight fragment of kerogen⁶. The two materials, which have almost identical elemental compositions¹³, undergo thermal fragmentation with equal readiness, yielding remarkably similar distributions of petroleum-like hydrocarbons^{14,15}, including the steranes and triterpanes¹⁶. Moreover, an activation energy reported¹⁷ for As decomposition, 49 kcal mol⁻¹, is well within the range of activation energies

reported¹⁸ for marine kerogen decomposition (46–56 kcal mol⁻¹). We shall assume, therefore, that the kinetics for As decomposition are similar to those for kerogen decomposition ($k_K \approx k_{As}$). To assess the stability of As relative to the hydrocarbons it generates, tridecylcyclohexane (TDCH) was also studied. When TDCH decomposes, carbon-carbon bond cleavage proceeds exclusively along the open chain¹⁹, thus giving a simple product distribution which is useful in comparative kinetic studies. Moreover, the light hydrocarbons generated from As and TDCH in a mixture can be measured separately under identical experimental conditions, so that the relative rates of decomposition can be measured in a common molecular environment.

Thermal decomposition of TDCH and As follows roughly first-order kinetics, with As decomposing about 10 times as fast as TDCH at 330 °C (Fig. 1). The ratio of their rate constants, k_{As}/k_{TDCH} , increases with decreasing temperature, following the linear Arrhenius relationship shown in Fig. 2. Projected to catagenic temperatures, k_{As}/k_{TDCH} is 24,000 at 100 °C and 2,100 at 150 °C. These projections are consistent with published values: McNab *et al.*¹⁷ reported that the thermal decomposition of a heavy bitumen from the Athabaska tar sands has an activation

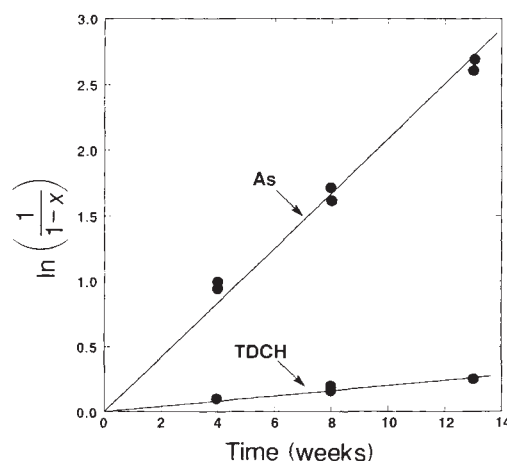


FIG. 1 Kinetic data for the thermal decomposition of TDCH in bitumen where x is the fraction of reactant converted after time t (weeks) at 330 °C. The extents of conversion were determined from three reactions carried out simultaneously: (1) a standard solution of TDCH (93.2 wt%) and adamantane (internal standard, 6.8 wt%); (2) a bitumen (As) from the Monterey formation, California Beta field, offshore Los Angeles. The crude sample (1.9° API gravity; 3.3% organosulphur) was first heated under argon for 48 hours at 100 °C to remove the light hydrocarbons; (3) a weighed mixture of standard solution (1) and As (2). All reactions were carried out in sealed glass tubes, and the products were analysed by gas chromatography (Hewlett Packard series 5880A, with flame ionization detector) on a 50-m column (Foxboro GB-5, internal diameter 0.25 mm, 0.1- μ m film) programmed at 3° min⁻¹ from 100 °C. Peak intensities were converted to wt% with respect to the (Ad) internal standard. Light hydrocarbon products (LHC) were arbitrarily taken to be the sum of all hydrocarbons eluting before adamantane (LHC < Ad). To estimate the rates of decomposition, n -dodecane (n -C₁₂) and the isoprenoid 2,6-dimethylundecane (C₁₃IP) in (3) were taken to indicate LHC production from As and from TDCH, using the observation that n -C₁₂ is generated from both As and TDCH whereas C₁₃IP is generated only from As. Assuming that As generates the same relative amounts of n -C₁₂ and C₁₃IP in (2) as it does in (3), the amount of n -C₁₂ generated from As in (3) is obtained from the product of (n -C₁₂/C₁₃IP) in (2) and the amount of C₁₃IP in (3). The excess n -C₁₂ in (3) is attributed to TDCH decomposition. It is further assumed that TDCH yields the same relative amounts of n -C₁₂ and (LHC < Ad)_{TDCH} in (1) as in (3). Thus the amount of LHC generated from TDCH in (3) is obtained from the product of (LHC < Ad)_{TDCH}/ n -C₁₂) in (1) and the amount of n -C₁₂ in (3) generated from TDCH. (LHC < Ad)_{As} in (3) is obtained by difference. Maximum conversion of As to (LHC < Ad)_{As} taken as 18.5% from the best fit to the kinetic data. Duplicate samples of (1), (2) and (3) were run at the indicated times, and at least two gas chromatograms were run on each product to minimize errors due to possible nonlinearity of the gas chromatogram system.

energy (49 kcal mol^{-1}) similar to that reported here (52 kcal mol^{-1}), and the rate of rate constants calculated from the kinetic equation of McNab *et al.*¹⁷ and that reported for the decomposition of *n*-heptane¹² is $k_{\text{As}}/k_{\text{Heptane}} = 23,800$ at 150°C .

The large differences in thermal stabilities reflected in $k_{\text{As}}/k_{\text{TDCH}}$ and $k_{\text{As}}/k_{\text{Heptane}}$ indicate that the carbon-carbon sigma bonds in hydrocarbons are very much stronger than the bonds that must be broken in As decomposition, and, by implication¹³⁻¹⁶, in kerogen decomposition. Hydrocarbons, therefore, should not decompose significantly under the conditions at which kerogen thermally decomposes to hydrocarbons. Moreover, petroleum hydrocarbons, irrespective of the relative stabilities of their precursors, should remain stable over geological time at temperatures below 150°C . The projected half-life for TDCH, for example, is 6×10^9 years at 150°C . These results are consistent with other kinetic studies where half-life approximations¹⁸ for the various hydrocarbons in petroleum range from tens of millions to billions of years at 150°C , leading some investigators²⁰ to propose that higher-molecular-weight hydrocarbons decompose to gas only at higher temperatures (between 150 and 190°C).

I turn now to the second analytical perspective on hydrocarbon stability: is there evidence in deposits of oil and gas for progressive thermal decomposition? Thermal decomposition of a mixture of hydrocarbons will concentrate the more stable compounds in the residual mixture. If the original mixture contains cycloalkanes and acyclic alkanes, the cycloalkanes should increase in concentration as decomposition progresses, because the thermal stability of carbocyclic compounds is very much higher than that of their open-chain homologues¹⁰. A mixture of cyclopentanes and isoheptanes, in which the ratio $[(1,1\text{-} + 1,3\text{-dimethylcyclopentanes} + \text{methylhexanes})/\text{dimethylpentanes}] = R$, provides a test for thermal decomposition. As the ratio R remains constant during petroleum generation²¹, any thermal decomposition of isoheptanes and cyclopentanes following oil generation should enrich the residual mixture in the more stable cyclopentanes, thereby shifting R outside the limits of range of values found in newly generated oil. A prepared mixture of these compounds, decomposing progressively at 390°C , shows the predicted concentration of cyclopentanes over time, with R increasing from an initial value of 4.1 to a value

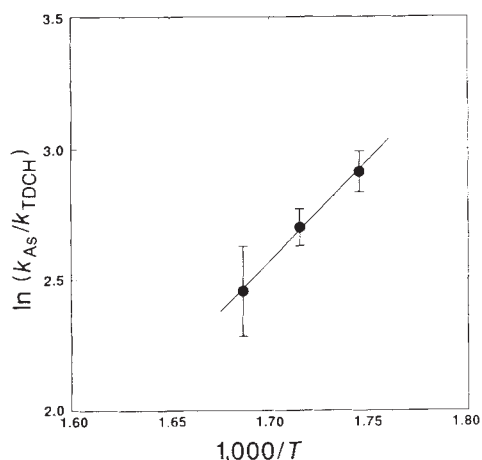


FIG. 2 Arrhenius plot of $\ln(k_{\text{As}}/k_{\text{TDCH}})$ against $1/T$, where T is in kelvin. The rate constants are for the thermal decompositions of As and TDCH in mixtures of the two, using the procedure described in the legend of Fig. 1. Reaction time at each temperature was selected for $\sim 50\%$ conversion of As. The mean As conversion was 51.2% for 112 days at 300°C , 56.9% for 56 days at 310°C and 54.5% for 26.5 days at 320°C . The error bars show $\pm 1\sigma$ for $\ln(k_{\text{As}}/k_{\text{TDCH}})$: six data points were obtained at 320°C , two data points each at 310 and 300°C . Linear regression of the means gives $\ln(k_{\text{As}}/k_{\text{TDCH}}) = 7.659/T - 10.44$, where $\ln k_{\text{TDCH}} = -33,738/T + 52.87$; $\ln k_{\text{As}} = -26,100/T + 42.5$, where T is in kelvin and k is in weeks.

of 21 at 86% decomposition. These hydrocarbons should respond similarly in a geological environment. For example, the C_7 hydrocarbons in large gas deposits (the so-called gas condensates), which are proposed to be the residue of thermal decomposition of oil to gas^{7,8}, should show increasing values of R as the extent of thermal decomposition approaches 100% . I could find little evidence for this, however, in a comprehensive analysis of oil and gas deposits. Only one sample, a condensate from a deep reservoir ($6,000 \text{ m}$) in the Lockridge field of the Permian Basin, Texas (Ordovician), had a ratio that indicated considerable thermal decomposition ($R = 17.8$). The R distribution representing the great majority of oils and condensates remains within the limits for oil generation, distinctly removed from those indicating partial decomposition (Fig. 3). Typical gas condensates show no evidence of residual debris from thermal decomposition. For example, a gas condensate from a gas field in Louisiana (Irene field, Upper Cretaceous Woodbine formation) shows $R = 4.17$, which is a common value for deep ($5,400 \text{ m}$) gas deposits in the Gulf of Mexico.

Deposits of oil and gas showing evidence of progressive thermal decomposition of oil to gas are statistically insignificant in the Earth. Thus, the thermal decomposition of hydrocarbons cannot be an important characteristic of petroleum habitats. This does not mean, however, that hydrocarbons do not decompose in the earth. Hydrocarbons at sufficient depth undergo oxidative decomposition by thermochemical reduction of sulphate²², and the products evidently are of altered composition. Moreover, the carbonaceous residues found with Palaeozoic gas deposits at great depth²³ can reasonably be expected to catalyse the decomposition of otherwise stable hydrocarbons. But my

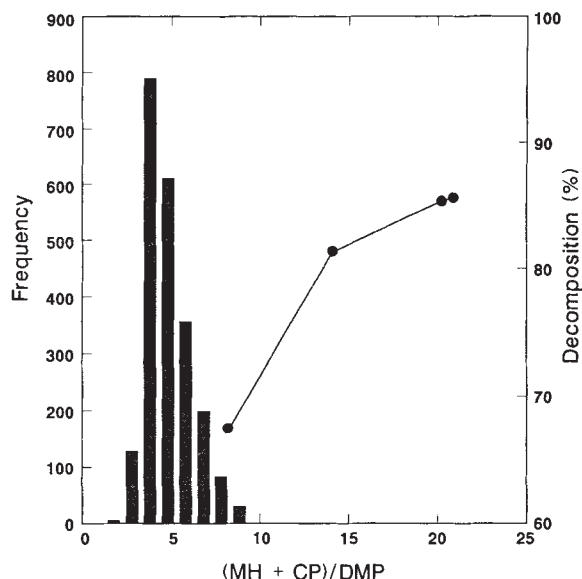


FIG. 3 Histogram of more than 2,000 petroleum samples. MH = 2-methylhexane + 3-methylhexane; CP = 1,1- + 1,3-dimethylcyclopentane (*cis* + *trans*); DMP = 2,4- + 2,3-dimethylpentane. This set does not contain samples believed to be biodegraded or to have suffered thermochemical sulphate reduction²². Primary oils from the Spraberry formation (Permian), Midland Basin, Texas, and the Saratoga Chalk formation (Upper Cretaceous), Sabine Parish, Los Angeles, define the outer limits of R : mean $R = ((\text{MH} + \text{CP})/\text{DMP}) \pm 1\sigma$, Midland 7.85 ± 0.382 , Sabine 3.23 ± 0.245 . All values of R between the Sabine and Midland limits are attributed to homologous sets of primary oils²¹, and thus represent invariant distributions of dimethylcyclopentanes and isoheptanes fixed at the time of oil generation, unaltered by post-generation processes. A prepared mixture of isoheptanes and dimethylcyclopentane (19.8% 2,3- + 2,4-dimethylpentanes, 46.93% 2- + 3-methylhexanes, 33.2% 1,1-dimethylcyclopentane) was heated in sealed glass at 390°C for up to 11 days. The solid line and points show the extent of thermal decomposition at 390°C .

results indicate that these processes can make no significant contribution to the pathways through which the light hydrocarbons are generated.

These conclusions support the hypothesis that the light hydrocarbons in petroleum are not generated by thermal decomposition of heavier hydrocarbons, but are instead the products of catalysis in kerogenous sedimentary rocks^{21,24}. □

Received 30 January; accepted 14 May 1991.

- Landes, K. K. *Bull. Am. Ass. Petrol. Geol.* **51**, 828–841 (1967).
- Colombo, U. in *Aspects of Petroleum Geochemistry* (eds Nagy, B. & Colombo, U.) 331–369 (Elsevier, New York, 1967).
- Evans, C. R., Rogers, M. A. & Bailey, N. J. L. *Chem. Geol.* **8**, 147–170 (1971).
- Hunt, J. M. *Petroleum Geochemistry and Geology* (ed. Gilluly, J.) 119–185 (Freeman, San Francisco, 1979).
- Waples, D. W. *Colorado Sch. Mines Quarterly* **78**, 15–30 (1983).
- Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence*, 69–267 (Springer, New York, 1984).
- Takach, N. E., Barker, C. & Kemp, M. K. *Bull. Am. Ass. Petrol. Geol.* **77**, 322–333 (1987).
- Barker, C. *Bull. Am. Ass. Petrol. Geol.* **74**, 1254–1261 (1990).
- Price, L. C. *J. Petrol. Geol.* **6**, 5–38 (1983).
- Mango, F. D. *Geochim. cosmochim. Acta* **54**, 23–27 (1990).
- Wall, L. A. *Nat. Bur. Stand. spec. Publ.* **357**, 47–60 (1972).
- Appleby, W. G., Avery, W. H. & Meerbott, W. K. *J. Am. chem. Soc.* **69**, 2279–2285 (1947).
- Orr, W. L. *Org. Geochem.* **10**, 499–516 (1985).
- Bandurski, E. *Energy Sources* **6**, 47–66 (1982).
- Pelet, R., Behar, F. & Monin, J. C. *Org. Geochem.* **10**, 481–498 (1985).
- Cassani, F. & Eglinton, G. *Chem. Geol.* **56**, 167–183 (1986).
- McNab, J. G., Smith, P. V. & Betts, R. L. *Ind. Engng. Chem.* **44**, 2556–2563 (1952).
- Ungerer, P. *Org. Geochem.* **16**, 1–25 (1990).
- Mushrush, G. W. & Hazlett, R. N. *Ind. Engng. Chem.* **23**, 288–294 (1984).
- Quigley, T. M. & Mackenzie, A. S. *Nature* **333**, 549–552 (1988).
- Mango, F. D. *Geochim. cosmochim. Acta* **54**, 1315–1323 (1990).
- Orr, W. L. *Bull. Am. Ass. Petrol. Geol.* **50**, 2295–2318 (1974).
- Raaben, V. F., Kalinin, N. A., Galimova, L. V. & Petrikevich, E. N. *Int. Geol. Rev. (USA)* **17**, 1317–1321 (1975).
- Mango, F. D. *Science* **273**, 514–517 (1987).

ACKNOWLEDGEMENTS. I thank Shell Development Company for their continuing support of this work and P. O. Knigge, H. I. Halpern and R. I. McNeil for support and advice.

Extraction of high-resolution carbonate data for palaeoclimate reconstruction

Larry A. Mayer

Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada

TEMPORAL variations in the calcium carbonate content of deep-sea sediments provide direct stratigraphic as well as important palaeoenvironmental information relating to the global carbon cycle. Here I present an algorithm that allows carbonate content and porosity to be accurately predicted from saturated bulk density in equatorial pelagic carbonates. Applying the algorithm to continuous laboratory measurements of density made on DSDP and ODP cores yields a nearly continuous carbonate record for the upper ~200 m of the sediment section. Long, ultra-high-resolution carbonate curves of this type should yield new insight into the evolution of the carbon chemistry of the oceans, as well as the role of external (Milankovitch) forcing in the development of the carbonate system. The algorithm can also be applied to quantitative, high-resolution seismic data, thereby enabling detailed carbonate records to be extracted from remotely derived geophysical data.

The record of calcium carbonate content in pelagic sediments is an extremely valuable stratigraphic and palaeoenvironmental tool. This is particularly true for the equatorial Pacific where down-section variations in carbonate content have been related to bio- and magneto-stratigraphies¹. Pleistocene glacial-interglacial climatic cycles are clearly recognizable in these records, as are longer-term trends^{2–4}. Deep-sea carbonates represent both a source and a sink in the global carbon dioxide cycle; the

record of their accumulation provides important insight into the history of global ocean chemistry and productivity^{5–7} and may allow determination of past levels of atmospheric carbon dioxide⁸.

Traditionally, the calcium carbonate content in marine sediments is determined by pressure calcimeter or coulometric techniques^{9,10}. These measurements are made on discrete samples of cored material. They have varying degrees of accuracy, but all are time-consuming and destructive to the sample, and few extremely high-resolution carbonate records (typically representing short intervals of time) have been produced.

Presented here is a technique that allows us to derive carbonate content from saturated bulk density (hereafter referred to as density) and possibly to extract data on carbonate content from high-resolution seismic records. The method has been developed for a two-component system (biogenic carbonate and silica), where changes in density result from marked differences in both the packing and grain density of the siliceous and carbonate components,¹¹ and thus the equations presented here are valid only for this system. Recent analyses of sediments from a three-component system (biogenic carbonate, silica and clay) reveal, however, that packing differences between the carbonate and silica/clay end members may result in considerable density differences and thus the method can still produce accurate predictions of carbonate content, although a different calibration is required. When applied to the continuous measurements of saturated bulk density standardly run on Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) cores, this technique allows rapid, non-destructive, nearly continuous measurement of calcium carbonate content. The bulk densities are measured by bore-hole logging and by gamma-ray attenuation. When coupled with seismic inversion techniques, the same approach provides a means of remotely extracting carbonate content (and thus palaeoenvironmental) information from digital seismic data.

To predict carbonate content from density, an empirical relationship between the two properties is established for near-surface samples and then corrections applied to adjust the density values of deeper samples to near-surface conditions.

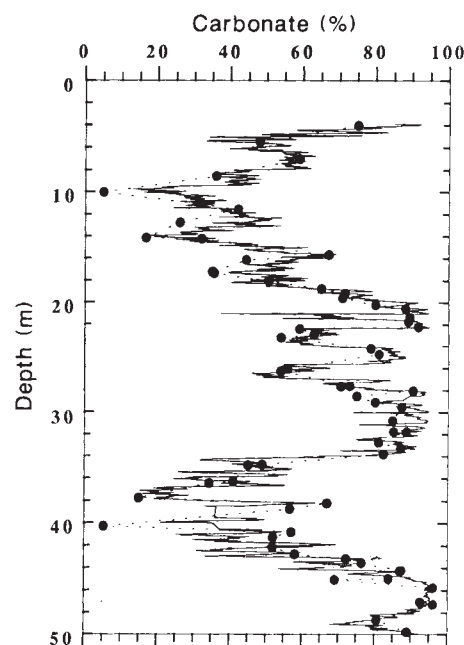
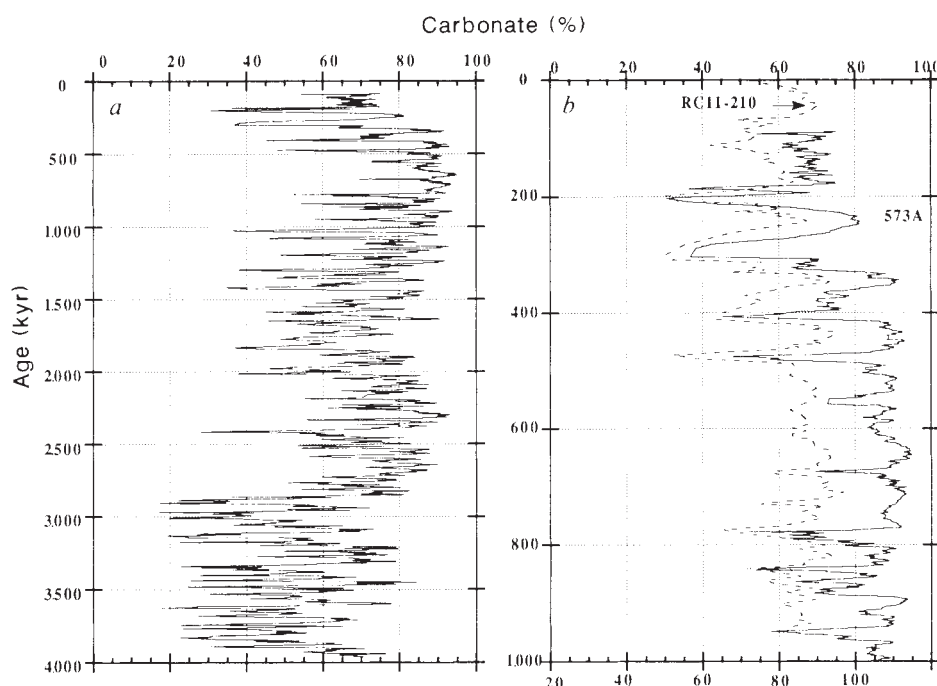


FIG. 1 Comparison of carbonate content predicted from gamma-ray attenuation data (solid line) with direct measurements of carbonate content made by coulometer in the laboratory (dots with dashed line) for upper 50 m of DSDP site 575B.

FIG. 2 *a*, Predicted carbonate stratigraphy from gamma-ray attenuation data for spliced record from DSDP sites 573 and 573A. Sample interval varies from ~500–1,000 years. Age model for upper 1 Myr is based on comparison of carbonate stratigraphy with core RC11-210, 300 km to the west^{25,26}. Stratigraphy for 1–4 Myr is based on palaeomagnetic dates²⁷. *b*, Comparison of predicted carbonate stratigraphy for upper 1 Myr with carbonate stratigraphy of piston core RC11-210. Note offset in carbonate scales.



Density measurements derived from samples whose depths are more than a few metres below the sea floor must be corrected for the effects of porosity rebound and compaction. Porosity rebound (the expansion of the sample when no longer subjected to overburden) can be estimated from the results of consolidation tests on deep-sea carbonates. A generalized relationship between porosity rebound, ϵ , and depth, z , for deep-sea carbonates has been established¹²:

$$\epsilon(\%) = 19.382z - 17.152z^2 \quad (1)$$

where z is in km. In the absence of direct measurements of porosity, density, ρ , can be related to porosity through the simple empirically determined linear relationship¹¹

$$\rho = 2.7074 - 0.173\phi$$

where ρ is in g cm^{-3} and ϕ is the porosity in %. The rebound-corrected densities should represent *in situ* values. We are interested only in the variation in density that results from changes in the relative proportions of the two components of the system (that is, the palaeoenvironmental signal) and therefore must also remove the change in density that results from compaction of the sediment column. Compaction can be represented by the generalized relationship between depth and density for deep-sea carbonates¹²:

$$\rho = 1.393 + 1.631z - 1.373z^2$$

(z in km) where the intercept value represents the average density of the surface samples in the database. The compaction signal is removed from the density curve, and density is then transformed to carbonate content by an empirically derived relationship for near-surface samples¹¹:

$$\% \text{ carbonate} = -835.52 + 1112.69\rho - 332.54\rho^2$$

This second-order relationship results from a linear grain-density factor (~10% of the signal) and a non-linear packing factor (the effect of spiny siliceous microfossils).

Applying the above relationships to the near-continuous measurements of density by gamma-ray attenuation on DSDP Leg 85 (ref. 13) results in an excellent match ($\pm 4\%$) between predicted carbonate content and that measured on discrete samples. The predicted record, however, provides a much more detailed picture of the evolution of the oceanic carbonate system (Fig. 1). The predictive relationship seems to hold, down to

depths of ~150–200 m below sea floor where the generalized compaction curve breaks down because of, amongst other factors, the onset of diagenetic alteration. The deviation of the predicted carbonate curve from actual measurements is, in its own right, an extremely sensitive indicator of the onset of diagenesis.

Given the sample interval of the gamma-ray attenuation data (~1 cm), the resulting carbonate curve has a temporal resolution of 500–1,000 years (depending on sedimentation rate). This level

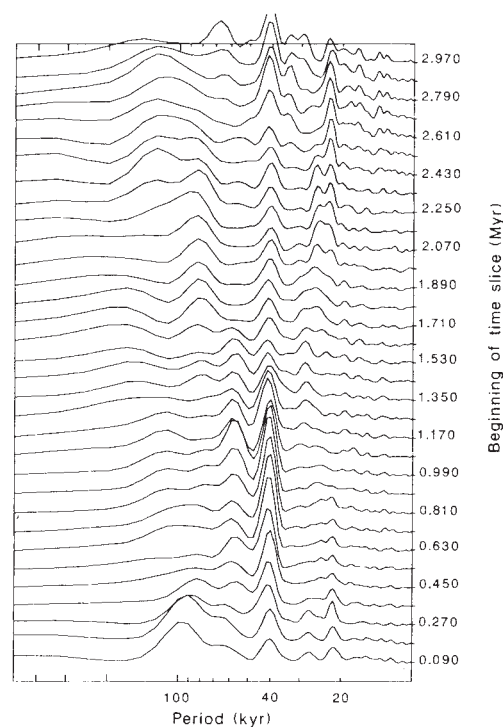


FIG. 3 Evolutionary spectra for spliced DSDP site 573 carbonate record. The first spectrum represents the time interval from 90,000 to 990,000 yr BP. Each successive spectrum represents an offset of 90,000 yr (90% overlap) and a 900,000-yr window. The age axis is keyed to the beginning of each window. The spectra are variance-conserving.

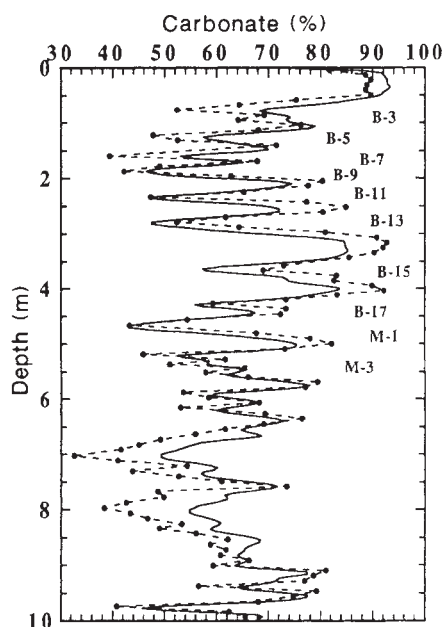


FIG. 4 Comparison of carbonate content predicted from the inversion of a Deep-Tow 4-kHz seismic record (solid line) with carbonate content measured on a core taken at the site of the seismic profile (dots with dashed line). Notations B-3 to M-3 represent well-dated equatorial Pacific carbonate events¹ (B-3=0.11 Myr; B-5=0.18 Myr; B-7=0.235 Myr; B-9=0.310 Myr; B-11=0.390 Myr; B-13=0.460 Myr; B-15=0.585 Myr; M-1=0.745 Myr; M-3=0.840 Myr) thus demonstrating the ability to extract direct stratigraphic as well as carbonate data from the seismic record.

of resolution, in combination with duplicate coring, allows us to resolve a number of stratigraphic ambiguities normally associated with the high-resolution drilling record (core-break gaps and missing sections) and to explore the role of external forcing on the carbonate system at a scale previously unattainable. Ultra-high-resolution carbonate curves from two holes of a single DSDP site (573) were carefully compared and spliced together to produce a continuous composite record for the past four million years (Fig. 2). The resulting high-resolution time series allows us to generate evolutionary spectra and to observe in detail the temporal variation in the ocean's response to orbital forcing (Fig. 3).

Limitations of the stratigraphic resolution create ambiguities in the evolutionary spectra: are variations real or the result of interpolation between a limited number of age-depth control points? The continuous high-resolution nature of the predicted carbonate record, however, permits the 'tuning' of the record and can be used to produce extremely accurate timescales¹⁴. In addition, the long, high-resolution time series produced allows bispectral techniques to be used to look at nonlinear interactions in climate forcing¹⁵.

The evolutionary spectra clearly reveal the previously documented¹⁶⁻²⁰ transition from a dominance of 41-kyr (obliquity) periodicity before ~900 kyr BP (before present) to a dominance of 100-kyr periodicity after that time. Moreover, the 41-kyr response increases in strength back to ~1.6 Myr, after which there is an interval (~1.6-2.45 Myr) of low strength at all frequencies which coincides with a period of weak solar forcing²¹. Beyond ~2.45 Myr the precessional frequencies gain strength and in the oldest part of the record (~3.1-3.8 Myr) the 41-kyr periodicity again increases in strength. These spectra represent the response of the carbonate system (ocean chemistry), and yet their temporal variability closely matches that of a long record of oxygen isotope ratios in the equatorial Pacific (site 677)²², implying a link between ice-volume changes and ocean chemistry. These extremely long and high-resolution carbonate records allow us to examine, in great detail, the response of one component of the ocean-atmosphere system to orbital forcing. Comparing these records (from diverse geographic settings) with records representing other components of the ocean-atmosphere system, such as oxygen and carbon isotopes, and applying more quantitative spectral techniques (for example, complex demodulation¹⁶), we can begin to piece together a detailed picture of the ocean's response to climatic forcing.

Finally, the ability to predict carbonate content from density values allows us to predict carbonate content remotely, from the seismic record. We have applied an autoregressive impedance inversion²³ to a Deep-Tow seismic profile from the central equatorial Pacific. Impedance variations in this environment are controlled almost entirely by density variations¹¹, and thus from the inverted impedance we can predict carbonate content (Fig. 4). Although this technique is limited by the resolution of the seismic source, recent advances in high-resolution profiling²⁴ make possible the remote mapping of carbonate variations on the scale of tens of centimetres in the upper hundred metres of the sediment column. We therefore have the opportunity to map carbonate fluxes over both space and time in large areas of the ocean basins. □

Received 21 December 1990; accepted 5 April 1991.

- Hays, J. D., Saito, T., Opdyke, N. D. & Burckle, L. H. *Geol. Soc. Am. Bull.* **80**, 1481-1514 (1969).
- Arrhenius, G. *Rep. Swedish Deep-Sea Exped. 1947-1948* **5**, 1-228 (1952).
- Berger, W. H. *J. for. Res.* **3**, 187-195 (1973).
- Adelsbeck, C. G. & Anderson, T. F. *Geology* **6**, 388-391 (1978).
- Arrhenius, G. *Paleogeogr. Paleoclimatol. Paleocool.* **67**, 119-146 (1988).
- Berger, W. H. & Keir, R. S. in *Climate Processes and Climate Sensitivity* AGU Geophys. Mag. Monogr. Ser. **29** (eds Hansen, J. & Takahashi, T.) 337-351 (Washington DC, 1984).
- Pedersen, T. F. *Geology* **11**, 16-19 (1983).
- Berger, W. H. & Spitz, A. *Paleoceanography*, **3**, 401-411 (1988).
- Jones, G. A. & Kaiteris, P. *J. sedim. Petrol.* **53**, 655-660 (1983).
- Chaney, R. C., Slonim, S. M. & Slonim, S. S. in *Geotechnical Properties Behavior and Performance of Calcareous Soils*, ASTM STP 777 (eds Demars, K. R. & Chaney, R. C.) 3-15 (Am. Soc. Testing and Materials, Philadelphia, 1982).
- Mayer, L. A. *J. sedim. Petrol.* **49**, 819-836 (1979).
- Hamilton, E. L. *J. sedim. Petrol.* **46**, 280-300 (1976).
- Wilkins, R. H. & Handyside, T. in *Initial Rep. DSDP 85* (eds Mayer, L.A. & Theyer, F.) 839-848 (US Govt Printing Office, Washington, DC, 1985).
- Herbert, T. D. *EOS* **71**, 1378 (1990).
- Hagelberg, T. K., Pisias, N. & Elgar, S. *EOS* **71**, 1378 (1990).
- Pisias, N. G. & Moore, T. C. *Earth planet. Sci. Lett.* **52**, 450-458 (1981).
- Moore, T. C., Pisias, N. G. & Dunn, D. *Marine Geol.* **46**, 217-233 (1982).
- Prell, W. in *Initial Rep. DSDP 68* (eds Prell, W. L. & Gardiner, J. V.) 455-464 (US Govt Printing Office, Washington, DC, 1982).
- Shackleton, N. J., Imbrie, J. & Pisias, N. G. *Phil. Trans. R. Soc.* **B318**, 679-688 (1988).
- Ruddiman, W. F. & Raymo, M. E. *Phil. Trans. R. Soc.* **B318**, 411-430 (1988).
- Berger, A. *Quat. Int.* **2**, 1-14 (1989).
- Shackleton, N. J. & Hall, M. A. in *Proc. Ocean Drill. Program 111* (eds Becker, K. & Sakai, H.) (US Govt Printing Office, Washington DC, in the press).
- Oldenburg, D. W., Scheuer, T. & Levy, S. *Geophysics* **48**, 1318-1337 (1983).
- Schock, S. G., LeBlond, L. R. & Mayer, L. A. *Geophysics* **54**, 445-450 (1989).
- Chuey, J. M., Rea, D. K. & Pisias, N. G. *Quat. Res.* **28**, 323-339 (1987).
- Pisias, N. G. & Rea, D. K. *Paleoceanography* **3**, 21-37 (1988).
- Weinreich, N. & Theyer, F. *Initial Rep. DSDP 85* (eds Mayer, L.A. & Theyer, F.) 849-904 (US Govt Printing Office, Washington, DC, 1985).

ACKNOWLEDGEMENTS. This research was supported by the NSERC (Canada), the US Office of Naval Research and the CNRS, France. Physical property data on Leg 85 was collected by R. Wilkins and T. Handyside. I thank N. Pisias, David Rea, Y. Lancelot and E. Vincent for their advice.

Extreme sexual dimorphism in a Miocene hominoid

Jay Kelley* & Xu Qinghua†

Division of Biology and Medicine, Brown University, Box G-B, Providence, Rhode Island 02912, USA

* Present address: Department of Oral Anatomy, College of Dentistry, The University of Illinois at Chicago, 801 South Paulina, Chicago, Illinois 60680, USA

† Institute of Vertebrate Paleontology and Paleoanthropology, Academia Sinica, PO Box 643, Beijing 100044, China

SOME Miocene hominoids may have been extremely sexually dimorphic for body size, inferred from the apparent dimorphism of dental and gnathic remains¹⁻⁴. But this has never been demonstrated convincingly for any fossil species because of small sample sizes, uncertainties about the number of species in most fossil samples, and the inability to reliably sex individual specimens. Here we demonstrate a case of extreme dental dimorphism, and presumed body-size dimorphism, in a Miocene hominoid sample in which these limitations have been overcome. *Lufengpithecus lufengensis* from the late Miocene site of Lufeng, China, was more dimorphic than the most dimorphic living hominoid, the orangutan, and may have been more dimorphic than any living anthropoid.

L. lufengensis is known from a large sample of over 1,000 specimens from a single outcrop. Included are several crania and many complete and partial jaws with dentitions or isolated but associated teeth belonging to single individuals⁵. The sample was originally described as comprising two species, one large and one small^{6,7}, a taxonomy adopted for subsequent analyses of sexual dimorphism^{8,9}. But analysis of metric variability in the large dental samples shows that in a large-and-small two-species alternative, each of the species would have numerous measures of metric variability substantially below the minimum

TABLE 1 Indices of dental dimorphism in *Lufengpithecus* and *Pongo*

	P ₄	M ₁	M ₂	M ₃	M _{avg}	PC _{avg}
<i>Lufengpithecus</i>	1.35	1.40	1.40	1.39	1.40	1.39
N (M, F)	10, 10	9, 11	10, 9	6, 9		
<i>Pongo</i> sampling $\bar{x}$	1.24	1.20	1.25	1.28	1.24	1.24
N (M and F ranges)	8-10	8-10	8-10	6-10		
s.d.					0.066	0.067
Sampling range					1.14-1.34	1.13-1.35

The measure of postcanine tooth-size used was tooth area (mean mesiodistal length $\times$ mean buccolingual breadth); only the lower dentition was used because there were more associated lower teeth for *Lufengpithecus*. The index of sexual dimorphism was calculated as male mean area/female mean area. Averages were calculated from the individual tooth positions for the whole molar row (M_{avg}) and for the entire masticatory postcanine (P_4 - M_3) dentition (PC_{avg}); individual tooth positions were not analysed separately because they are not independent. For *L. lufengensis*, indices for left and right sides were calculated separately and then averaged; the maximum difference between sides was 2.2%. In cases where one antimer was missing, the remaining tooth was used for both left and right side calculations. Sample sizes (N) for male (M) and female (F). For *P. pygmaeus*, we created a probability distribution of values for the index of dimorphism at each tooth position through repetitive sampling (with replacement) of a database of dental measurements including 41 males and 39 females, selecting sample sizes roughly equal to those of the Lufeng samples of sexed specimens. The database included only Bornean orang-utans measured by one individual²⁴. Sampling $\bar{x}$: means of 20 random samples of 10 males and 10 females each generated using a subroutine of the Systat statistical package. N (M and F ranges): ranges of sample sizes for each sex for the 20 samples. Sampling range: range of values for the 20 samples. Variability in sample sizes for both *Pongo* and *Lufengpithecus* results because not every specimen preserves the entire tooth row from P₄-M₃. The high values for *L. lufengensis* could reflect at least in part a greater male:female differential in interproximal wear (differentially affecting mesiodistal length) than in *Pongo*. But analysis of buccolingual breadth separately, which is affected by only the most pronounced wear, produced similar results. Also, changes in body size over time in a less dimorphic species might possibly produce similarly elevated measures of dimorphism in limited samples drawn from the entire timespan represented. Although this would be difficult to test without precise stratigraphic control (not available for Lufeng), there is no evidence of such a time effect for individual tooth positions, where it might be expected to produce platykurtic distributions for linear measurements (Q. Xu, *et al.*, manuscript in preparation).

values and lower 95% confidence intervals of extant apes¹⁰. By contrast, a single-species hypothesis cannot be rejected by these same criteria^{10,11} nor by dental morphological criteria³³. Moreover, the pattern of morphological differences that distinguishes large and small crania in the Lufeng sample is the same as that which distinguishes males and females in all extant great apes and in a variety of extant, sexually dimorphic cercopithecoid monkeys^{5,12}. Two-species alternatives in which the two species broadly overlap in size are also unlikely given the patterns of metric distribution of the teeth, particularly the canines and premolars (Q. Xu *et al.*, manuscript in preparation).

Although we cannot be certain that only one species is being sampled, it is most likely, given the sample sizes and completeness of the material available for study; the lack of evidence for the presence of more than one species; the metric arguments against the presence of large and small species; and the conformity with expectations in cranial morphology for the presence of a single species.

We determined the degree of sexual dimorphism in the post-canine dentition, to be used as an indicator of body-size dimorphism. Indirect measures of dimorphism, such as overall species variability¹³, do not produce estimates with sufficient

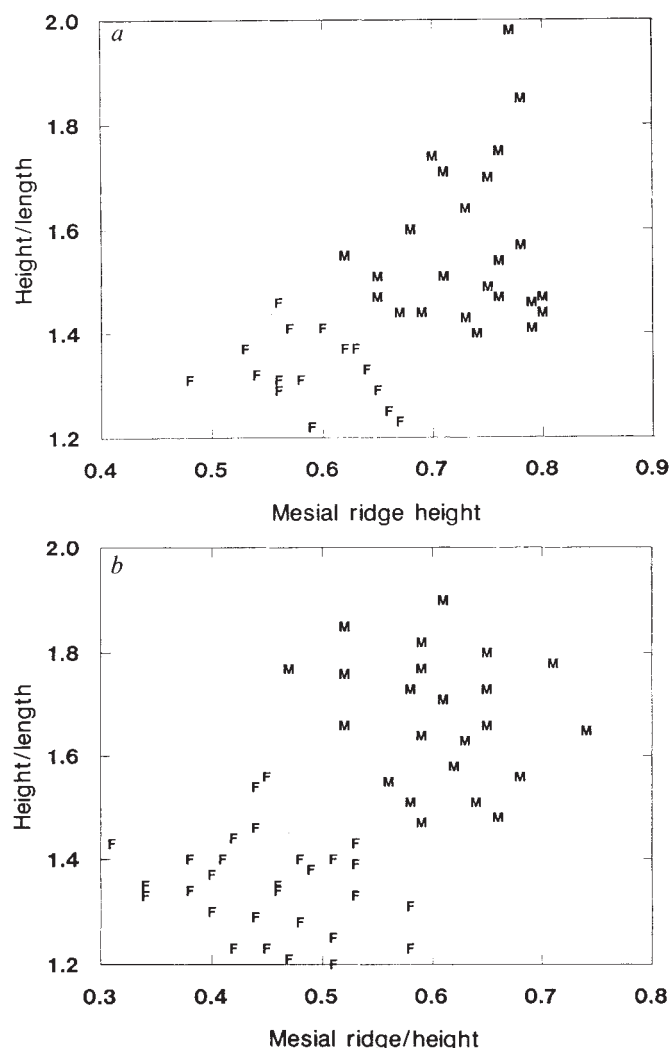


FIG. 1 Bivariate plots of lower canine shape indices for *Gorilla gorilla* (a) *P. pygmaeus* (b); M, male individuals; F, female individuals; height/length, maximum buccal crown height/maximum mesiodistal length, mesial ridge/height, mesial ridge length (measured from the crown apex to the mesial corner of the lingual cingulum)/maximum buccal height. These two species exemplify the consistent shape differences between male and female canines evident in all living great apes (J.K., manuscript in preparation).

precision. This requires a direct measure, which in turn requires a means to sex individual specimens. Because postcanine tooth-size itself is unreliable for determining the sex of specimens, jaws and associated dentitions were determined to be male or female using the canine teeth and, for lower dentitions if the canine was missing, the third premolar (P_3). Thus, only postcanine teeth associated with canines or P_3 can be used in the

calculation of sexual dimorphism. The canines were sexed using methods described previously^{1,10}, and were then used to aid in the sexing of P_3 s (Figs 1 and 2).

Indices of postcanine dental dimorphism were calculated for *L. lufengensis* and compared with the same indices for the orang-utan, *Pongo pygmaeus*. In the postcanine dentition, *Pongo* is the most dimorphic living hominoid¹⁴⁻¹⁶, and it is among the most sexually dimorphic of all living anthropoids, as a species perhaps the most dimorphic¹¹. Only baboons show a roughly equivalent degree of dimorphism¹⁶ (Fig. 3), with individual populations occasionally having somewhat greater levels of dimorphism than the *Pongo* sample used here¹⁷. Our aim was to determine the probability that dental dimorphism in *Lufengpithecus* was greater than in the most dimorphic living hominoid, the null hypothesis being that *L. lufengensis* was no more dimorphic than *P. pygmaeus*.

Sexual dimorphism in the postcanine dentition of *L. lufengensis* exceeds that of the orang-utan (Table 1). Two average values of dimorphism for the postcanine teeth of *Lufengpithecus* lie outside the ranges of values produced through repeated sampling of a database of *Pongo* dental measurements. They lie more than two standard deviations beyond the means of the *Pongo* samples.

We thus conclude that *L. lufengensis* was more sexually dimorphic in the postcanine dentition than the most dimorphic living

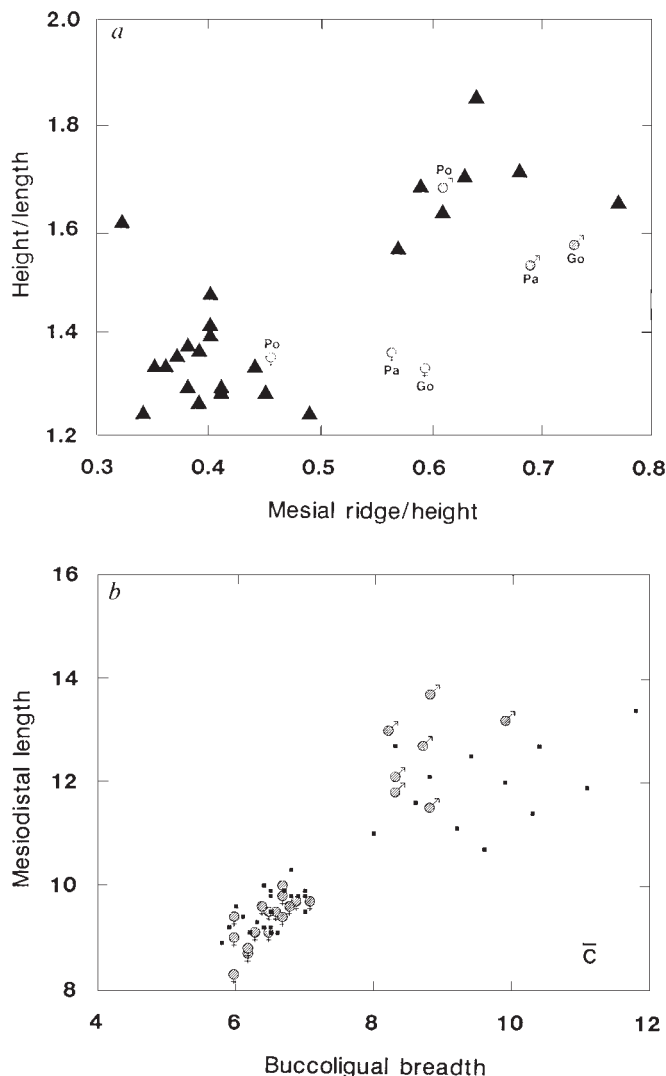


FIG. 2 *a*, Bivariate plot of lower canine shape indices for Lufeng specimens ($\blacktriangle$), plus male and female means of *G. gorilla* (Go), *P. pygmaeus* (Po) and *Pan troglodytes* (Pa). This plot includes all of the complete, unworn or minimally worn Lufeng lower canines, either isolated or associated with postcanine dentitions. By these shape parameters, the Lufeng canines fall into distinct male and female groups. *b*, Bivariate plot of Lufeng lower canine mesiodistal length against buccolingual breadth for the entire lower canine sample. Male/female symbols indicate specimens sexed according to shape indices and shown in *a*. This plot shows that the two size groups of Lufeng lower canines clearly define male and female groups. Because male and female canines do not overlap in size in this species, even partial canines that cannot be characterized according to their shape (solid squares) can be confidently sexed, as can the teeth with which they are associated. Thus we sexed 20 partial or complete lower dentitions (11 female, 9 male). We also sexed 8 additional lower dentitions (5 female, 3 male) using the third premolar. Lower third premolars from Lufeng ($N=36$) also fall into two nonoverlapping size groups. All of those known to be male based on associated canines fall into the large-sized group and all of those known to be female fall into the small-sized group. Thus, lower third premolars and the teeth associated with them can also be confidently sexed based only on size. Apparent difference from *a* in number of sexed females results from two specimens having identical crown dimensions.

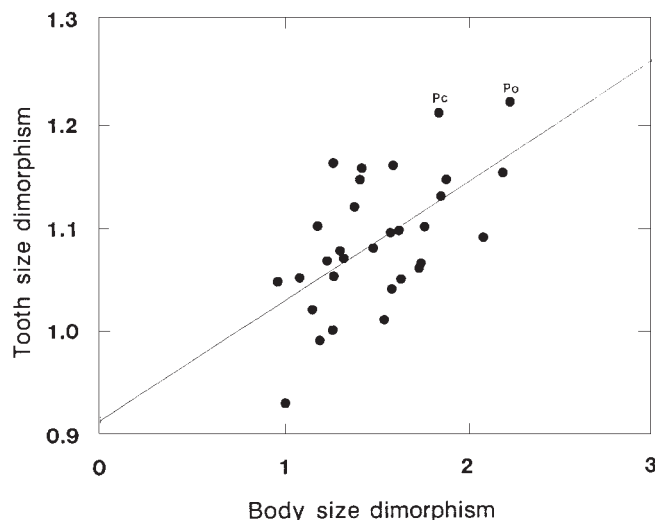


FIG. 3 Regression of mandibular postcanine tooth-size dimorphism against body-size dimorphism. Tooth-size dimorphism equals PC_{avg} described in Table 1; body-size dimorphism calculated as male weight/female weight. Tooth dimorphism = $0.914 + 0.115$ (body size dimorphism); Kendall Tau-B, 0.411 ($P < 0.001$). The high correlation shown here differs from the results of an earlier study²⁵ in which no correlation was found between the two variables. Our reexamination of the previous study, however, in light of more recent and more accurate body-size data²⁶⁻²⁸ revealed that body-weight dimorphism had been substantially underestimated in several species with large postcanine dental dimorphism. Although the correlation here is highly significant, only 35% of the variance in tooth-size dimorphism is explained by body-size dimorphism, making a precise estimate of body-size dimorphism in *Lufengpithecus* unreliable. Rough estimates based on molar size indicate an average weight around 85% of the 63 kg average of *P. pygmaeus*, and a dimorphism index of about 2.6 (a direct estimate of body size and body-size dimorphism is not yet possible as only four, mostly fragmentary postcranial bones of *L. lufengensis* have been recovered⁵). Dental data^{16,24,29-32} and body-weight data²⁶⁻²⁸ are from published sources. Slightly lower value of tooth-size dimorphism for *Pongo* than in Table 1 results from the use of male/female means for the entire *Pongo* sample rather than the sampling means used for Table 1. Po: *P. pygmaeus*; Pc: *Papio cynocephalus*.

hominoid and, probably, any living anthropoid. An examination of the relationship between postcanine tooth-size dimorphism and body-size dimorphism in 31 anthropoid species indicates that the extreme dental dimorphism in *L. lufengensis* may reflect an equally great and perhaps unprecedented degree of body-size dimorphism. (Fig. 3.)

Adopting a single-species taxonomy for Lufeng, these results show that the greatest sexual dimorphism in tooth-size, and probably body-size, found among extant anthropoid primates does not represent some upper limit during primate evolutionary history. This has important implications for the taxonomic interpretation of other, less well-represented fossil hominoid

samples. The presence of seemingly unacceptably high levels of sexual dimorphism in fossil samples that are morphologically homogeneous does not necessarily require the recognition of more than one species. Patterns and degrees of sexual dimorphism not represented among extant primates may have characterized many fossil hominoid species¹⁸. We have only a limited understanding of how developmental processes (bimaturism, growth rates)¹⁹, ecological pressures (feeding competition, female metabolic requirements)^{20,21} and behavioural demands (mating system)^{22,23} interact to produce particular patterns of sexual dimorphism, and different patterns from those known should not be unexpected. □

Received 13 November 1990; accepted 29 April 1991.

1. Kelley, J. *J. hum. Evol.* **15**, 461–495 (1986).
2. Kelley, J. & Pilbeam, D. in *Comparative Primate Biology*, Vol. 1 (eds Swindler, D. R. & Erwin, J.) 361–411 (Liss, New York, 1986).
3. Pilbeam, D. *Am. Anthropol.* **88**, 295–312 (1986).
4. Pickford, M. *Hum. Evol.* **1**, 111–148 (1986).
5. Wu, R., Xu, Q. & Lu, Q. *Acta anthrop. Sinica* **5**, 1–30 (1986).
6. Xu, Q., Lu, Q., Pan, Y., Zhang, X. & Zheng, L. *Xexue Tongbao* **9**, 544–556 (1978).
7. Xu, Q. & Lu, Q. *Vert. Palasiat.* **17**, 1–13 (1979).
8. Wu, R. & Oxnard, C. E. *Nature* **306**, 286–260 (1983).
9. Wu, R. & Oxnard, C. E. *Am. J. Primatol.* **5**, 303–344 (1983).
10. Kelley, J. & Etler, D. *Am. J. Primatol.* **18**, 15–34 (1989).
11. Plavcan, J. M. thesis, Duke Univ. (1990).
12. Wu, R., Xu, Q. & Lu, Q. *Acta anthrop. Sinica* **2**, 1–10 (1983).
13. Kay, R. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 209–212 (1982).
14. Martin, L. B. thesis, Univ. London (1983).
15. Oxnard, C. E., Lieberman, S. S. & Gelvin, B. R. *Am. J. Primatol.* **8**, 127–152 (1985).
16. Swindler, D. R. *Dentition of Living Primates* (Academic, New York, 1976).
17. Swindler, D. R., McCoy, H. A. & Hornbeck, P. V. in *The Baboon in Medical Research* Vol. 2 (ed. Vagtbord, H.) 133–150 (University of Texas Press, Austin, 1967).
18. Kelley, J. thesis, Yale Univ. (1986).
19. Shea, B. T. *Hum. Evol.* **1**, 97–110 (1986).
20. Demment, M. W. *Afr. J. Ecol.* **21**, 219–233 (1983).

21. Pickford, M. *Hum. Evol.* **1**, 77–90 (1986).
22. Jarman, P. *Biol. Rev.* **58**, 485–520 (1983).
23. Kay, R. F., Plavcan, J. M., Glander, K. E. & Wright, P. C. *Am. J. phys. Anthropol.* **77**, 385–397 (1988).
24. Mahler, P. E. thesis, Univ. Michigan (1973).
25. Post, D., Goldstein, S. & Melnick, D. *Am. J. phys. Anthropol.* **49**, 533–544 (1978).
26. Leutenegger, W. & Cheverud, J. *Int. J. Primatol.* **3**, 387–402 (1982).
27. Jungers, W. L. in *Size and Scaling in Primate Biology* (ed. Jungers, W. L.) 345–382 (Plenum, New York, 1985).
28. Fleagle, J. G. *Primate Adaptation and Evolution* (Academic New York, 1988).
29. Leutenegger, W. *Am. J. phys. Anthropol.* **35**, 91–100 (1971).
30. Lucas, P. W., Corlett, R. T. & Luke, D. A. *Z. Morph. Anthropol.* **76**, 253–276 (1986).
31. Fleagle, J. G. & Kitahara-Frisch, J. in *The Lesser Apes* (eds Preuschoft, H., Chivers, D. J., Brockelman, W. Y. & Creel, N.) 192–206 (Edinburgh University Press, 1984).
32. Pilbeam, D. *Bull. Peabody Mus. nat. Hist.* **31**, 1–185 (1969).
33. Wood, B. A. & Xu, Q. *J. hum. Evol.* **20**, 291–311 (1991).

ACKNOWLEDGEMENTS. We thank R. Smith, D. Pilbeam, L. Martin, G. Albrecht, J. Miller and L. Flynn for comments and discussions, the Erikson Biographical Institute for computer facilities, the Departments of Mammalogy at the British Museum (Natural History), the United States National Museum and the Museum of Comparative Zoology, Harvard University for access to primate specimens in their care. This research was supported by the National Academy of Sciences Foreign Scholar Exchange Program, the Chinese Academy of Sciences and the L. S. B. Leakey Foundation.

Fitness consequences of foraging behaviour in the zebra finch

William C. Lemon*

Department of Zoology, University of Texas, Austin, Texas 78712, USA

OPTIMAL foraging theory is based on the assumption that natural selection favours animals that forage most efficiently^{1–3}. But such selection does not act directly on foraging efficiency, but rather indirectly by favouring animals that survive and reproduce most successfully. Studies that use optimal foraging models often assume that maximization of some behavioural currency, such as the animal's net rate of energy gain, maximizes the animal's fitness⁴, but rarely is an attempt made to test this assumption^{5–8}. Most studies of the effects of foraging behaviour on fitness fail to control for the amount of energy gained by the foraging animals^{5–8}, and lead to the obvious conclusion that animals that eat more reproduce more. Often the studies do not control for characters correlated with foraging behaviour⁶ or compare traits assumed to be correlated with fitness^{7,8}. A better method would be to assign net rates of energy gain to randomly chosen individuals for their entire lifetimes in a controlled environment and measure fitness directly. Variation in the amount of energy consumed would be controlled by using individuals that employ a time-minimizing foraging strategy⁹ and would alter the time taken to satisfy their daily energy requirements, while obtaining the same absolute amount of energy. I have now manipulated the net rate of energy gain in four populations of the zebra finch *Taeniopygia guttata*, and show that fitness, as measured by population growth rate, is

indeed positively and significantly correlated with the net rate of energy gain.

Four populations of zebra finches were housed in identical indoor aviaries (2.0 × 3.0 × 2.5 m) at constant temperature (27 °C) and day length (14 h). Each population initially contained 15 pairs of randomly assigned adult zebra finches (about 90 days old). Offspring of these finches were moved to a separate aviary two weeks after fledging. Each aviary contained 30 nest sites. Nest material, oyster shell grit, cuttle bone and water were provided *ad libitum*. Each population received 200 g of mixed

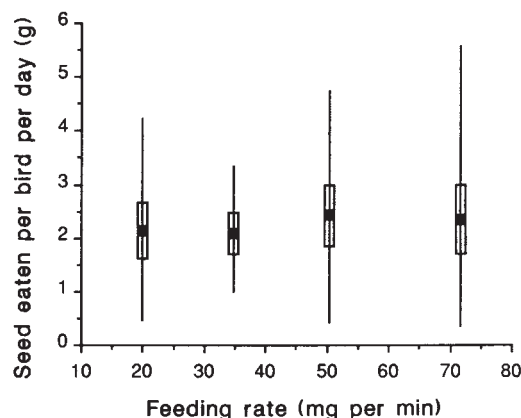


FIG. 1 Mass of seed eaten per adult bird per day by zebra finches in experimental populations with feeding treatments limiting the maximum feeding rate. Means, standard deviations, and ranges are shown. The mass of seed eaten per adult bird per day in each population was interpolated from linear regressions of the amount of seed eaten per bird in the population versus the population size, which provided an estimate of how much seed adult birds ate while controlling for the number of offspring being supported.

* Present address: Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

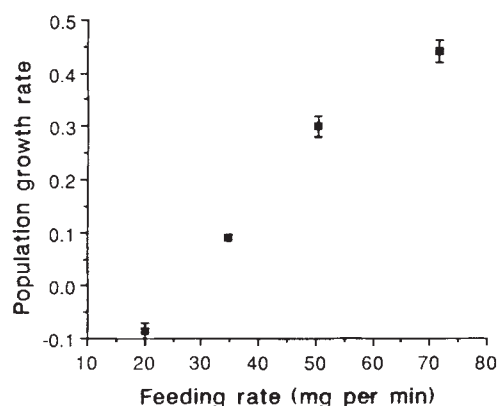


FIG. 2 Population growth rates of four populations of zebra finches with different maximum feeding rates, calculated for three-month intervals. The exponential population growth rates and jackknife estimates of the 95% confidence intervals¹⁸ were calculated from the dominant eigenvalues of the population projection matrices¹⁶.

grass seeds daily, which was more than they could consume in a day, placed in the bottom of a 25 × 50 × 30 cm glass box. They also received either 20 g of boiled egg or 20 g of peas on alternate days.

The populations were treated identically except for the manner in which their seed was presented. The control population (group A) received only 200 g of seed daily, while the three experimental populations received 200 g of seed daily mixed with 400 g (group B), 600 g (group C) or 800 g (group D) of empty seed hulls. Foraging zebra finches remove the hulls before ingesting the kernels. The birds in the experimental treatments were forced to search through the hulls to locate whole seeds. These experimental treatments decreased the maximum feeding rate of the birds in those populations without altering the amount of food available.

Maximum feeding rates were estimated from 35 individually housed adult zebra finches during their first foraging bout of the day in a pre-weighed patch of seed. Foraging time was measured by an observer behind a blind. Each bird was subjected to all four treatments in random order on four consecutive days. The maximum feeding rates ($\text{mg min}^{-1} \pm 95\%$ confidence intervals) were: A, 71.6 ± 7.7 ; B, 50.3 ± 5.8 ; C, 34.6 ± 10.0 ; D, 20.0 ± 6.2 . Maximum feeding rates were significantly different

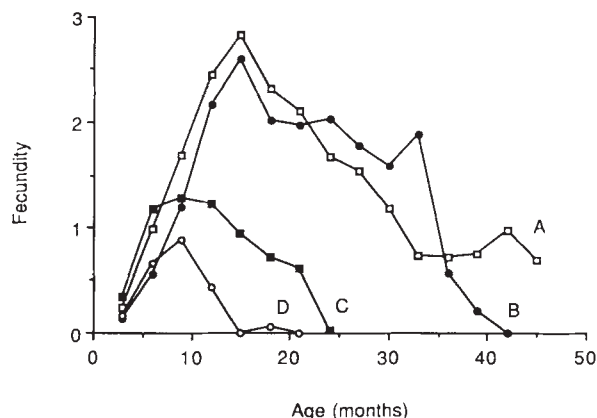


FIG. 3 Fecundity of four populations of zebra finches with four different maximum feeding rates. Age-specific fecundity is the number of female offspring produced per adult female per three-month interval adjusted for the probability of surviving through the interval¹⁶. A census of the breeding populations was performed daily to determine reproductive success and mortality of offspring from each adult female.

(repeated measures analysis of variance (ANOVA), $F = 26.10$, degrees of freedom (d.f.) = 3, 139, $P < 0.0001$; Fischer's protected least significant difference, $P < 0.05$ for all pair-wise comparisons), significantly correlated with the ratio of chaff:seed (simple regression, $F = 90.61$, d.f. = 1, 139, $P < 0.0001$), and typical of feeding rates seen in wild granivorous passerines¹⁰⁻¹².

Seed consumption was measured by providing each population with pre-weighed seed or seed and hulls daily, and weighing the remains 24 h later (Fig. 1). The mean amount of seed eaten per adult bird per day was not correlated with maximum feeding rate (Kendall's coefficient of rank correlation¹³ $\tau = 0.5$, $n = 4$, $P > 0.05$). Adult males and females housed individually and fed with the same four treatments consumed about the same amount of seed daily as the adults in the experimental populations, and there was no difference between the sexes (two-factor ANOVA: treatment, $F = 0.003$, d.f. = 3, 90, $P > 0.5$; sex, $F = 0.068$, d.f. = 1, 90, $P > 0.5$; interaction, $F = 0.234$, d.f. = 3, 90, $P > 0.5$). This does not preclude the possibility that males in the harsher treatments were excluding females from the feeders, but in more than 500 h of observation this was not observed.

To calculate the net rate of energy gain, I measured the increase in metabolic rate of adult birds foraging on various ratios of hulls to seed by the indirect volumetric method^{14,15}. The metabolic rates of foraging birds were not correlated with their feeding rates ($r = 0.077$, $n = 27$, $P > 0.05$). Regardless of the rate at which they obtained seeds, foraging birds expended energy at the same rate. Because the net rate of energy gain is the difference of the rate of energy gain and the rate of energy expense, the observed feeding rates were directly proportional to net rates of energy gain.

The experimental treatments significantly affected the rate at which birds could obtain food without systematically altering the amount of food that was obtained. This response was consistent with the prediction for animals using a time-minimizing

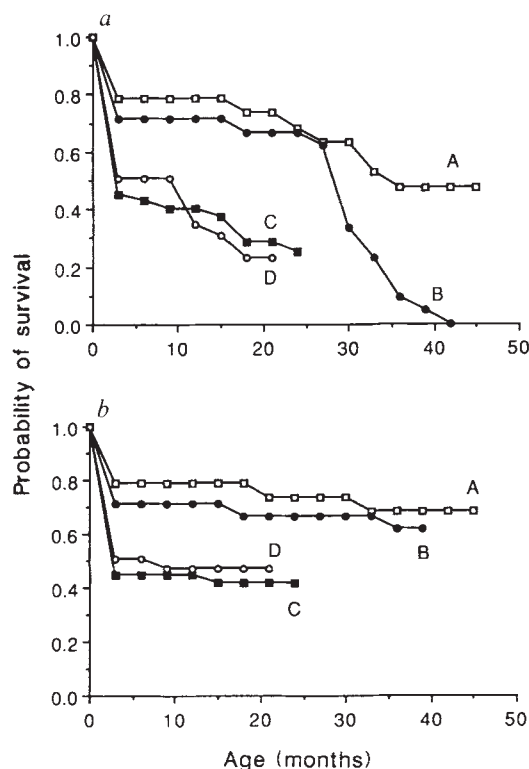


FIG. 4 Probability of survival from birth to age x of female (a) and male (b) zebra finches in experimental treatments that provided four different maximum feeding rates. Survival was determined by following populations until mortality of the original cohort of females exceeded 75%.

foraging strategy¹⁵. This experimental design was analogous to assigning a behavioural phenotype to each member of the population that limited the maximum feeding rate.

The exponential population growth rate, which was derived from age-specific reproduction and mortality of the individuals in each population, was used as an index of fitness^{16,17} (Fig. 2). I also compared two components of fitness: fecundity (Fig. 3) and the probability of survival (Fig. 4). Both fecundity averaged over time and female survival beyond 12 months of age were positively correlated with the maximum feeding rate (Kendall's coefficient of rank correlation $\tau=1.0$, $n=4$, $P<0.05$, one-tailed). The variances in lifetime reproductive success of the four populations were not significantly different (log-ANOVA test for homogeneity of variances¹³ $F=3.39$, d.f. = 3, 8, $P>0.05$), indicating that the experimental treatments had similar effects on all the individuals within each population. Population growth rate, and therefore fitness¹⁷, was positively correlated with maximum feeding rate (Kendall's coefficient of rank correlation $\tau=1.0$, $n=4$, $P<0.05$, one-tailed).

Studies of foraging birds that use a time and energy budget to predict fitness¹⁹ assume that the measured budget has a fitness consequence. The explicit assumption of long-term rate-maximization foraging models is that fitness increases monotonically as the net rate of energy gain while foraging increases⁴. In the study described here, I tested this assumption. I found

that fitness in zebra finches was positively correlated with the net rate of energy gain. Therefore, a common assumption of optimal foraging theory, that natural selection favours animals with a higher net rate of energy gain, seems to be valid. □

Received 19 November 1990; accepted 17 May 1991.

1. Krebs, J. R. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 22–63 (Blackwell, Oxford, 1978).
2. Pyke, G. H. A. *Rev. Ecol. Syst.* **15**, 523–575 (1984).
3. Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton Univ. Press, Princeton, 1986).
4. Schoener, T. W. A. *Rev. Ecol. Syst.* **2**, 396–404 (1971).
5. Wise, D. H. *Oecologia* **41**, 289–300 (1979).
6. Fritz, R. S. & Morse, D. H. *Oecologia* **65**, 194–200 (1985).
7. Ritchie, M. E. *Evol. Ecology* **2**, 232–252 (1988).
8. Ritchie, M. E. *Oecologia* **82**, 56–67 (1990).
9. Hixon, M. A. *Am. Nat.* **119**, 569–599 (1982).
10. Benkman, C. W. *Auk* **107**, 376–386 (1990).
11. Benkman, C. W. & Pulliam, H. R. *Ecology* **69**, 1195–1199 (1988).
12. Sullivan, K. A. *Anim. Behav.* **39**, 380–387 (1990).
13. Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, San Francisco, 1981).
14. Miller, B. G., Kirkwood, J. K., Howard, K., Tuddenham, A. & Webster, A. J. F. *Lab. Anim.* **51**, 313–317 (1981).
15. Hails, C. J. *Condor* **85**, 61–65 (1983).
16. Caswell, H. *Matrix Population Models* (Sinauer, Sunderland, Massachusetts, 1989).
17. Cooper, W. S. J. *theor. Biol.* **107**, 603–629 (1984).
18. Meyer, J. S., Ingersoll, C. G., McDonald, L. L. & Boyce, M. S. *Ecology* **67**, 1156–1166 (1987).
19. Morrison, M. L., Ralph, C. J., Verner, J. & Jehl, J. R. *Studies Avian Biol.* **13**, 1–515 (1990).

ACKNOWLEDGEMENTS. I thank L. Fleishman, C. Lively, A. Mills and J. Phillips for helpful comments on the manuscript.

Lekking in the black grouse—a test of male viability

Rauno V. Alatalo, Jacob Höglund* & Arne Lundberg*

Department of Biology, University of Jyväskylä, Yliopistonkatu 9, SF-40100, Jyväskylä, Finland

* Department of Zoology, Uppsala University, Box 561, S-75122 Uppsala, Sweden

LEKS, where males congregate to display and females attend only to mate, present one of the most remarkable outcomes of sexual selection¹. It is a common but untested belief that females mate with the most vigorous males². In leks of the black grouse *Tetrao tetrix*, males dominant in winter flocks were most successful in mating, as were males winning fights over female dummies placed at territory boundaries. Males tear feathers from each others' tail ornaments in combats, and attractive males always had undamaged tails. We report here that by choosing victorious males, females mate with males that are most likely to survive the following six months. There is a strong association between female preference and male viability which supports a basic assumption of the 'good gene' models (refs. 3–9) where choosy females benefit through better viability of their offspring. But females may also be choosing viable males for immediate benefits^{2,10}, in particular if diseases can be transmitted during copulations.

Black grouse males were caught at a feeding site in central Finland (Petäjävesi, 62°10' N, 20°05' E) during the winters of 1987–90, and marked individually with colour rings. Copulations were recorded on the four leks within 3 km of the site of capture, in late April and early May^{11,12} each year. Females typically copulate only once^{11–15} (79% of the 29 radiotagged females did so). Females move freely on the lek, visit several males on several mornings before final solicitation and copulation^{11,12}. Female preference is not related to male morphology, as wing, tail, tarsus, sternum sizes and body mass explain only 2.2% of variance in ln-transformed copulation number ($n=75$, $F=0.69$, $P=0.66$). Nor can mating success be predicted by straightforward measures of display activity; rates of vocalizations, tail ornament exposure and constancy of presence only accounted for 7.5% of variance ($n=61$, $F=1.70$, $P=0.12$).

The most preferred male on each lek spent more time fighting with neighbours (30.2% of his presence time, s.d. = 13.9, $n=9$ males) than the other territorial males (19.8%, s.d. = 14.4, $n=52$, U -test, $z=2.02$, $P<0.05$). In their territories, males circle visiting females^{12,13} and preferred males were able to keep neighbouring males farther away (8.8 m, s.d. = 5.4, $n=13$ males) compared with the other males (6.0 m, s.d. = 3.8, $n=46$, U -test, $z=2.01$, $P<0.01$). Additional support for the theory that good fighters are preferred is that males frequently tear feathers from each others' tail ornaments, in particular from the conspicuous white undertail. Among 31 males, 45% had visible tail damages, but the most attractive always had intact tails (Table 1).

To test directly if fighting ability is related to female choice, we performed two experiments. First, to overcome the problem that each male is dominant inside his own territory, we estimated the dominance relationship of yearling males at the feeding site just before their first lekking season. We left oats in a few small areas where only one male could forage at each moment, and checked all the dominance interactions between males. The few

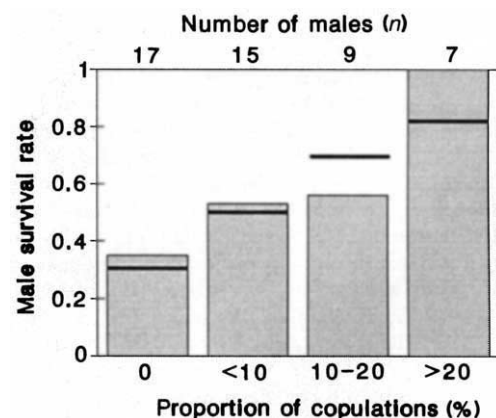


FIG. 1 Proportion of older males surviving until the next autumn against the proportion of copulations attained within the lek. If the same male survived between several seasons he was given, to assure statistical independency, the average seasonal mating success. The horizontal lines indicate the predictions from the logistic regression model (for slope: $t=2.52$, $P<0.05$, $n=48$).

TABLE 1 Female preference in relation to male dominance and damages of the tail ornament

	Male attractivity		Test (two-tailed)
	Low	High	
Relative dominance position in winter flocks (s.e.)*	0.26 ± 4	0.56 ± 13	U-test, $U=12$ $P<0.05$, $n=22$
Proportion winning a female dummy†	0.20	0.80	Binomial test $P<0.05$, $n=15$
Proportion with tail damages‡	0.58	0.00	Fisher exact $P<0.01$, $n=31$

* On the basis of interactions with at least five different males we estimated the proportion of all the males dominated by each yearling male. Preferred males ($n=4$) performed at least one copulation in the following lekking period.

† Female dummies were placed on the boundary between two older males' territories, and the male preferred by live females was classified as attractive. We used 15 different pairs of territorial males on six leks; the attractive males having, on average, 5.8 and the nonpreferred males 0.5 copulations. The three nonpreferred males taking over the dummy obtained more 'natural' copulations ($\bar{x}=1.7$) than the other non-preferred males ($\bar{x}=0.1$, U-test, $U=0.5$, $P<0.01$). In six control treatments the nonpreferred males always monopolized a dummy placed in the centre of their territories.

‡ Territorial older males were checked for missing feathers in the white undertail or in the lyre. Males with at least three copulations ($n=7$) were classified as attractive.

yearlings that later in the spring succeeded in copulating had been dominant in these feeding competition situations (Table 1).

Second, we placed a taxidermic dummy female exactly on the territorial boundary of two older males, at a site where males often fought. We hypothesized that because of the increased value of the territorial combat the dominant male would monopolize the dummy at the boundary. The male that was more attractive to real females exercising a free choice was usually the one to copulate with the dummy, often after severe fighting (Table 1). In the control cases, where dummies were inside territories of the less successful males, these males were always able to retain the dummy in their possession.

As males are unable to overtake females in other males' territories, females can choose between males, and thus the lekking system is not based on male coercion. Disturbance of copulations occurred in 16.7% of the 318 copulations witnessed, which may force the female to copulate once more. But, in seven cases of disturbance, a marked female copulated again with the same male in three cases and with another male in four cases, but never with the disturbing male. Females do not seem to seek undisturbed copulations, as they clearly preferred to mate on larger leks with more than 10 males¹², where disturbance occurs (13.4%, s.d. = 12.6, $n=11$ leks), whereas on smaller leks disturbance is very rare (1.8%, s.d. = 0.6, $n=11$ leks, U-test, $z=2.63$, $P<0.01$).

Male survival was checked 6 months later in October–January at the start of winter feeding. The main cause of mortality is predation by goshawks *Accipiter gentilis* and foxes *Vulpes vulpes*^{16,17}. Disappearing males must have died as males only disperse in their first year of life and subsequently stay in the same winter flock^{13,17}. We also visited leks 3–6 km from the site of capture. Altogether 17 of the 84 yearling males marked in the study flock were observed on those distant leks, but none of the 55 older males was ever observed outside the 3 km radius ($\chi^2=12.7$, $P<0.001$). Yearling males obtain few copulations^{12–14}, thus we used data on males that are at least 2 years old (Fig. 1).

The survival probability of the attractive males was more than twice that of those with no matings (Fig. 1). The average number of copulations per season for surviving males was 3.6 (range 0–14, $n=26$ individual males) and for nonsurviving males 1.0 (range 0–5, $n=22$, U-test, $z=2.63$, $P<0.01$). The result is not confused by age-dependent survival because, in a multiple logistic regression, survival was explained by number of copulations ($t=1.98$, $P<0.05$, $n=76$) but not by age ($t=-0.03$, $P>0.90$;

based on 23 local 1-year-old, 22 2-year-old and 31 >2-year-old males).

Male fighting ability, indicated by intact ornaments and by subtle differences in the dominance relationships of the territorial males, thus seem to indicate male viability. For nonlekking animals a few studies have demonstrated an association between size of male ornaments and viability^{9,18–20}. Together with this report, these data indicate that females could be gaining genes for viability, but for a final test it is necessary to study offspring viability which so far has only been done under laboratory conditions^{21,22}. Alternatively, some immediate benefits might explain female choosiness^{2,10}. In the black grouse, where risk of predation on the lek is minute¹², avoidance of disease transmission during copulation is a possible benefit not checked. □

Received 1 November 1990; accepted 29 May 1991.

- Bradbury, J. W. & Gibson, R. M. in *Mate Choice* (ed. Bateson, P. P. G.) 109–138 (Cambridge University Press, Cambridge, 1983).
- Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
- Zahavi, A. *J. theoret. Biol.* **53**, 205–214 (1975).
- Hamilton, W. D. & Zuk, M. *Science* **218**, 384–387 (1982).
- Pomiankowski, A. N. *Oxford Surv. evol. Biol.* **5**, 136–184 (1988).
- Grafen, A. *J. theoret. Biol.* **144**, 473–516 (1990).
- Grafen, A. *J. theoret. Biol.* **144**, 517–546 (1990).
- Iwasa, Y., Pomiankowski, A. & Nee, S. *Evolution* (in the press).
- Møller, A. P. *Proc. R. Soc. Lond. B243*, 145–148 (1991).
- Reynolds, J. D. & Gross, M. R. *Am. Nat.* **136**, 230–243 (1990).
- Höglund, J., Alatalo, R. V. & Lundberg, A. *Behaviour* **114**, 221–231 (1990).
- Alatalo, R. V., Höglund, J., Lundberg, A. & Sutherland, W. J. *Behav. Ecol.* (in the press).
- Koivisto, I. *Finnish Game Res.* **26**, 1–60 (1965).
- Kruijt, J. P. & de Vos, G. J. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 279–290 (University of Chicago Press, Chicago, 1988).
- Kruijt, J. P., de Vos, G. J. & Bossema, I. *Proc. Int. Ornithol. Congr.* **15**, 399–423 (1972).
- Angelstam, P. *Ornis Scand.* **15**, 123–134 (1984).
- Willebrand, T. thesis, Uppsala Univ. (1988).
- Watt, W. B., Carter, P. A. & Donohue, K. *Science* **233**, 1187–1190 (1986).
- von Schantz, T. et al. *Nature* **337**, 166–169 (1989).
- von Schantz, T., Göransson, G., Fröberg, I., Helgee, A. & Witzell, H. *Anim. Behav.* **40**, 89–104 (1990).
- Partridge, L. *Nature* **283**, 290–291 (1980).
- Taylor, C. E., Pereda, A. D. & Ferrari, J. A. *Am. Nat.* **129**, 721–729 (1987).

ACKNOWLEDGEMENTS. We thank L. Gustafsson, A. P. Møller, M. Petrie and S. Ulfstrand for comments, M. Halonen, H. Hakkarainen, M. Hovi, E. Huhta, J. Jokimäki, P. Rintamäki, O. Rätti, P. Siikamäki, J. Suhonen, J. Sundberg and S. Thorman for help in the field. L. Ijäs, P. Valkeajärvi for advice and P. Angelstam for the grouse traps. R.A. was supported by the Academy of Finland and J.H. and A.L. from the Swedish NSRC.

Depth is encoded in the visual cortex by a specialized receptive field structure

Gregory C. DeAngelis, Izumi Ohzawa & Ralph D. Freeman*

Bioengineering and Neurobiology Groups, School of Optometry, University of California, Berkeley, California 94720, USA

BINOCULAR neurons in the visual cortex are thought to perform the first stage of processing for the fine stereoscopic depth discrimination exhibited by animals with frontally located eyes. Because lateral separation of the eyes gives a slightly different view to each eye, there are small variations in position (disparities), mainly along the horizontal dimension, between corresponding features in the two retinal images. The visual system uses these disparities to gauge depth. We studied neurons in the cat's visual cortex to determine whether the visual system uses the anisotropy in the range of horizontal and vertical disparities. We report here that there is a corresponding anisotropy in the cortical representation of binocular information: receptive-field profiles for left and right eyes are matched for cells that are tuned to horizontal orientations

* To whom correspondence should be addressed.

of image contours. For neurons tuned to vertical orientations, left and right receptive fields are predominantly dissimilar. Therefore, a major modification is required of the conventional notion of disparity processing. The modified scheme allows a unified encoding of monocular form and binocular disparity information.

Horizontal disparities are thought to be encoded by a group of neurons with varying relative displacements (incongruities) of the receptive fields for the left and right eyes¹⁻⁵. The two receptive fields of these binocular neurons are also assumed to be identical in shape^{5,6}. This is shown schematically in Fig. 1a, where left and right receptive fields have the same luminance sensitivity profiles, but are located at noncorresponding (incongruous) points. The preferred disparity of such a cell is determined by how much the two fields are displaced from the point of retinal correspondence. On the basis of this, searches have been made for vertical-horizontal differences in the distribution of positional incongruities between left and right receptive fields. This difference is expected if the visual system encodes disparity information efficiently. Because the eyes are separated laterally, binocular parallax produces a much larger range of horizontal than vertical disparities. Thus, only a small range of vertical disparity values near zero requires encoding. A predominance of horizontal receptive-field incongruities was reported in the original study¹, but it has not been confirmed^{3,4,7,8}. The issue is complicated by the wide range of receptive-field eccentricities and the problem of eye-position monitoring^{4,9}.

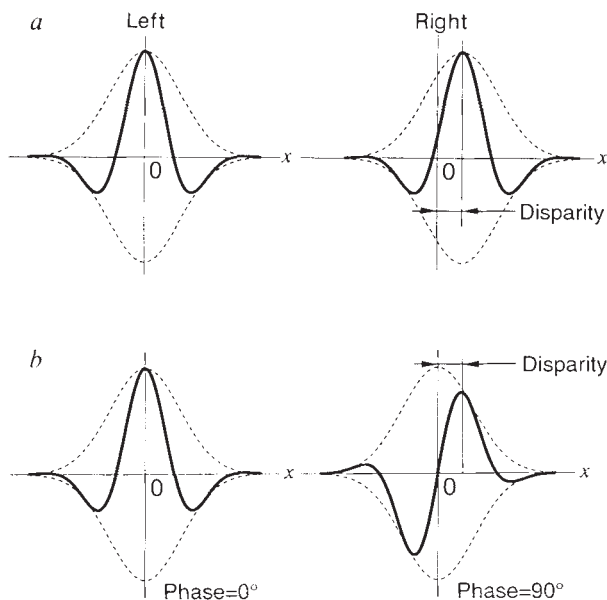


FIG. 1 Two possible schemes for disparity encoding by binocular simple cells. Simple cells⁶ have receptive fields consisting of spatially separate subregions which respond, alternately, to either bright stimuli (luminance increments) or dark stimuli (luminance decrements). The solid curves are Gabor functions which represent receptive-field profiles for the left and right eyes of a simple cell. The abscissa (x) for each curve gives retinal position and the ordinate represents sensitivity to a luminance increment, such that downward deflections of the curve correspond to dark-excitatory (or OFF) subregions and upward deflections correspond to bright-excitatory (or ON) subregions. The dashed curves show the Gaussian envelopes of the idealized receptive fields. *a*, In the conventional view, the two receptive-field profiles are identical in shape, but their centres (the centres of the Gaussian envelopes) are binocularly disparate (incongruous). Therefore, the cell is selective to a particular disparity by virtue of the spatial offset of the receptive fields from the point of retinal correspondence. *b*, An alternative scheme for disparity encoding, in which the left and right receptive-field profiles may differ in shape (or phase), but are centred at corresponding retinal locations. As a result, the optimal stimulus disparity is determined by the difference in phase between the two profiles, and by the size (or spatial frequency) of the fields.

An alternative to this traditional theory is possible if we do not assume that the profiles of the left and right receptive fields are identical (Fig. 1b). Suppose that the receptive-field centres, as defined by the peak of the envelope (dashed curve), are located at corresponding retinal points and the left and right receptive-field profiles differ in shape (as indicated by a 90° phase difference in Fig. 1b). Here the optimal disparity is determined by the difference in the shapes (or phases) of the receptive-field profiles, and by the size (or spatial frequency) of the field. This disparity encoding scheme links the preferred disparity of each neuron to its spatial-frequency selectivity^{4,10,11}, and integrates smoothly with quadrature encoding schemes for monocular image representation^{12,13}.

To distinguish which of the two mechanisms for the encoding of disparity actually operates, it is first necessary to compare the profiles of left and right receptive fields. Although previous reports claim that they are matched, the methods used were either qualitative⁶ or inadequate for firm conclusions⁵. We have now used appropriate quantitative methods to study cells in the striate cortex. From 17 adult cats, 131 simple cells were investigated and 61 were studied in sufficient detail for a complete analysis of both eyes. Standard surgical and physiological techniques were used (see legend to Fig. 2). For each isolated cell, preliminary tests were made with drifting grating patches to determine optimal values of orientation and spatial frequency for each eye. We also obtained binocular interaction profiles by varying the relative phase between gratings presented dichoptically¹⁴. Cells were classified as simple or complex according to standard criteria⁶ and also by the degree of modulation in the responses to gratings¹⁵.

To look at the receptive fields in more detail we used a version of the reverse correlation technique^{10,16}. This yields detailed two-dimensional receptive-field profiles for each eye. One-dimensional profiles are then obtained by integrating the two-dimensional data along the axis parallel to the cell's preferred orientation. To each one-dimensional profile, we fit a Gabor function^{12,17-19} (the product of a gaussian and a sinusoidal function) with five free parameters, one of which is the phase of the sinusoidal component relative to the centre of the gaussian envelope. The value of this phase parameter, Φ , determines the order and relative strengths of the bright- and dark-excitatory receptive-field subregions. The difference in phase, $\delta\Phi$, between the left and right profiles is then a measure of their similarity. For comparison across cells, the absolute value of phase difference, $|\delta\Phi|$, is used to quantify differences in shape between the left and right profiles. Values of $|\delta\Phi|$ close to 0° indicate that left and right profiles are closely matched. For values of $|\delta\Phi|$ near 180°, one profile is nearly inverted with respect to the other. Figure 2a shows typical results of the fitting process for three cells, all having preferred orientations close to vertical. Note the similarity of left and right profiles for the cell with $|\delta\Phi| = 13^\circ$. Almost completely opposite profiles are seen for the cell with $|\delta\Phi| = 159^\circ$, and a smaller mismatch is observed for a cell with $|\delta\Phi| = 113^\circ$.

Results for the 61 cells studied in detail are summarized in Fig. 2b. Here, $|\delta\Phi|$ is plotted against the preferred orientation for each neuron (horizontal = 0°, vertical = 90°). Note that virtually all cells with preferred orientations near horizontal have values of $|\delta\Phi|$ near 0°. In fact, 78% (14/18) of cells tuned to orientations within 20° of horizontal have $|\delta\Phi|$ values smaller than 30°. The left and right receptive-field profiles are well matched for these cells. By contrast, cells tuned to orientations within 20° of vertical have a wide variety of phase differences between profiles for the two eyes. The same results are obtained for orientation ranges of 5°, 10°, 20° or 30° around the extremes. These data provide evidence of a cortical specialization for processing vertical contours, which provide horizontal disparity information necessary for stereoscopic depth discrimination.

Our findings are not in accordance with some previous work. Hubel and Wiesel⁶ originally noted that left and right fields

were closely matched. But the authors mapped the fields qualitatively with bars of light and thus differences between the shapes of the left and right profiles might have been missed. Methodology was also a problem for Maske *et al.*,⁵ who reported also that the fields were closely matched. These authors obtained receptive-field profiles by recording peristimulus time his-

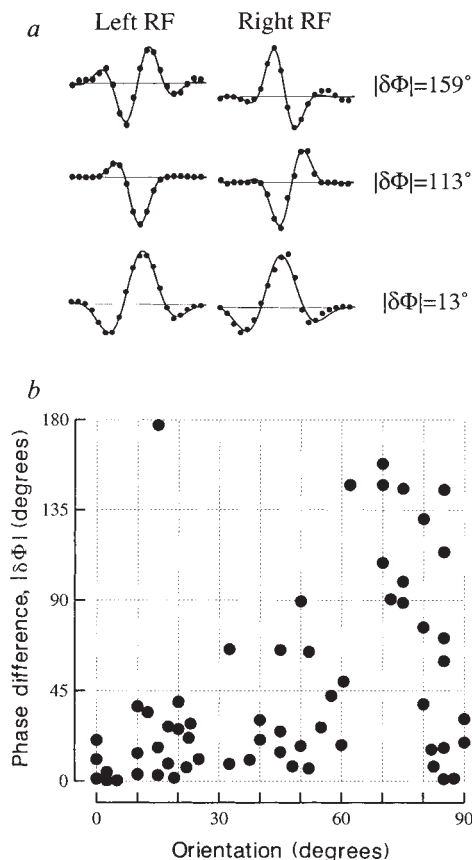


FIG. 2 A summary of receptive field mapping data for 61 simple cells. *a*, One-dimensional receptive-field profiles (●) and fitted Gabor functions (solid curves) are shown for the left and right eyes of three simple cells, whose preferred orientations are 70°, 85° and 85°, respectively (top to bottom). *b*, The absolute value of phase difference, $|\delta\Phi|$, between the left and right profiles is plotted against the preferred orientation for each cell. Orientation is shown as the number of degrees away from horizontal, such that 90° represents vertical.

METHODS. Animals were anaesthetized with halothane (3% in O_2) during surgery. During recording, cats were paralysed with gallamine triethiodide continuously infused at $10 \text{ mg kg}^{-1} \text{ h}^{-1}$. Sodium thiameylal was also infused at $1 \text{ mg kg}^{-1} \text{ h}^{-1}$. Artificial respiration used a gas mixture of N_2O , O_2 and CO_2 (70%:29%:1%). Action potentials from single neurons in striate cortex were recorded with 1 ms resolution. Visual stimuli were presented to either eye by means of a pair of cathode ray tube displays (mean luminance, 50 cd m^{-2}). Further details of these procedures are described elsewhere²⁵. Following standard tests with drifting gratings, the reverse correlation method¹⁶ was used to map the left and right receptive fields separately. In this test, we present successive 50 ms flashes of a small bar stimulus (usually $1.5^\circ \times 0.5^\circ$) at each of 20×20 locations on a stimulus grid which covers the receptive field. Both the location of the bar and whether it is bright or dark are randomized. Each time there is an action potential, a two-dimensional histogram is incremented at the grid location corresponding to the most likely causal stimulus. A composite receptive-field profile is formed by taking the difference of separate bright and dark histograms. The two-dimensional profile is then integrated along the axis of the cell's preferred orientation to produce a one-dimensional profile. In some cases, one-dimensional profiles were obtained directly by reverse correlation using long-bar stimuli. Each one-dimensional profile is approximated with a Gabor function having the form $K \exp(-(x-x_0)^2/a^2) \cos((x-x_0)/U_0 + \Phi)$, where K , x_0 , a , U_0 , and Φ are free parameters. The absolute value of the difference in Φ between the left and right receptive field profiles, $|\delta\Phi|$, is used as a measure of their similarity.

tograms in response to moving bright and dark bars. With this method, the shape of receptive-field profiles may be distorted by the effects of sequential stimulation of neighbouring subregions, that is, spatial and temporal response properties are confounded. In addition, a quantitative method was not used to compare left and right profiles, and it may have been difficult to discern phase differences of 90° or less (see, for example, Fig. 4 of ref. 5). This point is especially problematic in the study described in ref. 5, as data are presented for only 15 simple cells and orientation preferences are not given.

As we are proposing a major modification to the standard theory of the neural basis of stereoscopic depth discrimination, it is important to consider how this alternative scheme relates to the traditional types of disparity-tuning of cells in the visual cortex. The conventional view is represented most prominently by work on the behaving monkey for which four basic disparity response patterns have been identified: tuned-excitatory, tuned-inhibitory, near (or tuned-near) and far (or tuned-far)^{2,20}. A model has been proposed recently that incorporates the idea of mismatched left and right receptive fields to account for the

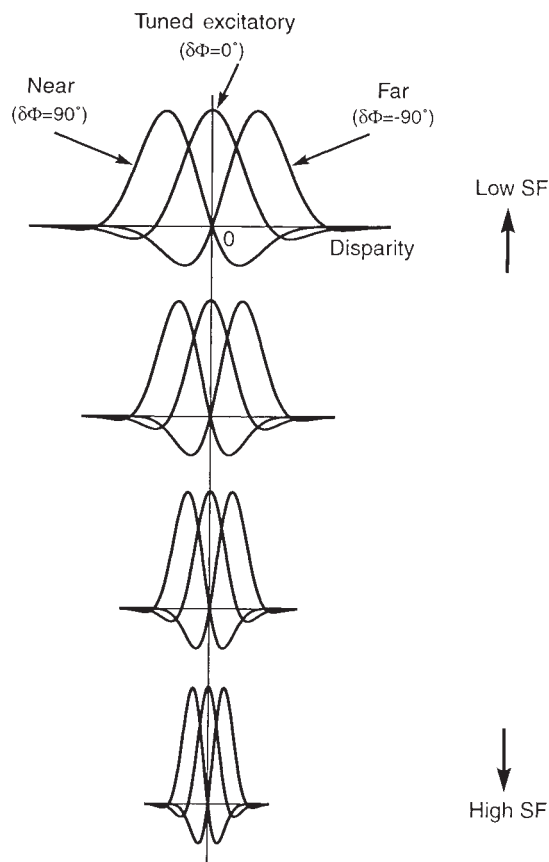


FIG. 3 A scheme for encoding a large range of binocular disparities through receptive-field phase. Disparity tuning curves are shown at each of four spatial frequency (SF) values ranging from low (top) to high (bottom). For each set of curves, the abscissa is binocular disparity (with zero disparity indicated by the vertical line), and the ordinate represents response strength. At each spatial frequency, there are simple cells which take on a broad range of values of $\delta\Phi$ (for clarity, only three examples are shown: $\delta\Phi = -90^\circ$, 0° and 90°). This produces many types of disparity tuning at each frequency scale. Some of these profiles are similar to the tuned-excitatory ($\delta\Phi=0^\circ$), near or tuned-near ($\delta\Phi=90^\circ$) and far or tuned-far ($\delta\Phi=-90^\circ$) types (solid curves), and others are intermediate categories. Neurons tuned to low spatial frequencies have large receptive fields and can encode a large range of disparities. Cells which are tuned to high spatial frequencies encode a much more limited range of disparities, but with higher precision. As a result of this multi-scale image analysis, the entire range of horizontal disparities to which the visual system is sensitive may be encoded.

basic response patterns²¹. As in our scheme (Fig. 1b), the simple cell in this model may have left and right receptive fields of different shape, but their centres are in retinal correspondence. When the receptive field phase difference, $\delta\Phi$, is zero, the model predicts a disparity-tuning function that is analogous to the tuned-excitatory response. When $\delta\Phi = 180^\circ$, the model produces a tuned-inhibitory type of disparity tuning. For other values of $\delta\Phi$, the predicted disparity tuning can be similar to that of a near cell ($\delta\Phi = 90^\circ$) or a far cell ($\delta\Phi = -90^\circ$). Therefore, the traditional types of disparity tuning are consistent with the alternative disparity-encoding scheme (Fig. 1b) provided that $\delta\Phi$ takes a broad range of values. We find that this requirement holds only for cells with preferred orientations close to vertical. Although our data suggest that horizontal disparity can be encoded solely by differences in receptive field phase, it should be noted that our findings do not rule out the possibility that positional disparities (or incongruities) are used in the neural processing of depth information. Positional incongruities between the left and right receptive field envelopes were not measured.

Finally, how does a population of neurons encode the entire range of horizontal disparity to which the visual system is sensitive? If simple cells encode disparity through receptive field phase, then the precision with which each neuron can signal disparity is commensurate with its optimal spatial frequency. We suggest that simple cells collectively encode the necessary range of binocular disparity through an arrangement which is illustrated schematically in Fig. 3. By having a variety of types of disparity tuning at each spatial frequency scale, it is possible to encode a large range of binocular disparity. Moreover, this scheme links the disparity sensitivity of cortical cells to their fundamental monocular receptive field characteristics, that is orientation, spatial frequency and phase. This scheme also predicts that the range of fine (high spatial frequency) disparity judgements should be smaller than that for coarse (low spatial frequency) disparities. Some psychophysical data are consistent with this prediction^{22,23}.

We have shown here that between the left and right eyes the receptive fields of cells tuned to orientations near vertical exhibit a broad range of phase differences. By contrast, cells tuned to orientations near horizontal tend to have matched receptive fields. This is consistent with simple cells encoding horizontal disparity through differences in the phases and not spatial displacement of the receptive fields. Thus, the striate cortex possesses a specialized mechanism for processing vertical contours which provide horizontal disparity information needed for estimating depth through stereopsis. We suggest that simple cells collectively encode the necessary range of binocular disparities through the distribution of their receptive-field phases and sizes (or spatial frequencies). This information, in turn, is transmitted to a population of complex cells which are ideally suited as disparity detectors²⁴. □

Received 7 February; accepted 1 May 1991.

- Barlow, H. B., Blakemore, C. & Pettigrew, J. D. *J. Physiol.* **193**, 327–342 (1967).
- Poggio, G. F. & Fischer, B. *J. Neurophysiol.* **40**, 1392–1405 (1977).
- von der Heydt, R., Adorjani, C. S., Hanny, P. & Baumgartner, G. *Expl Brain Res.* **31**, 523–545 (1978).
- Ferster, D. *J. Physiol.* **311**, 623–655 (1981).
- Maske, R., Yamane, S. & Bishop, P. O. *Vision Res.* **24**, 1921–1929 (1984).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
- Joshua, D. E. & Bishop, P. O. *Expl Brain Res.* **10**, 389–416 (1970).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **232**, 29–30 (1973).
- LeVay, S. & Voigt, T. *Visual Neurosci.* **1**, 395–414 (1988).
- Freeman, R. D. & Ohzawa, I. *Vision Res.* **30**, 1661–1676 (1990).
- Marr, D. & Poggio, T. *Proc. R. Soc. B* **204**, 301–328 (1979).
- Gabor, D. *J. Inst. electr. Eng.* **93**, 429–457 (1946).
- Sakitt, B. & Barlow, H. B. *Biol. Cybern.* **43**, 97–108 (1982).
- Freeman, R. D. & Robson, J. G. *Expl Brain Res.* **48**, 296–300 (1982).
- DeValois, R. L., Albrecht, D. G. & Thorell, L. G. *Vision Res.* **22**, 545–559 (1982).
- Jones, J. P. & Palmer, L. A. *J. Neurophysiol.* **58**, 1187–1211 (1987).
- Marcelja, S. *J. opt. Soc. Am.* **70**, 1297–1300 (1980).
- Daugman, J. G. *J. opt. Soc. Am.* **2**, 1160–1169 (1985).
- Jones, J. P. & Palmer, L. A. *J. Neurophysiol.* **58**, 1233–1258 (1987).
- Poggio, G. F., Gonzalez, F. & Krause, F. *J. Neurosci.* **8**, 4531–4550 (1988).
- Nomura, M., Matsumoto, G. & Fujiwara, S. *Biol. Cybern.* **63**, 237–242 (1990).

- Schor, C. M. & Wood, I. *Vision Res.* **23**, 1649–1654 (1983).
- Schor, C. M., Wood, I. C. & Ogawa, J. *Vision Res.* **24**, 573–578 (1984).
- Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. *Science* **249**, 1037–1041 (1990).
- Ohzawa, I. & Freeman, R. D. *J. Neurophysiol.* **56**, 221–242 (1986).

ACKNOWLEDGEMENTS. We thank G. Ghose for help with the experiments and for critical comments on the manuscript. The work was supported by the US National Eye Institute (research and CORE grants) and by the New Energy and Industrial Technology Development Organization of Japan.

Acceleration of activation and inactivation by the β subunit of the skeletal muscle calcium channel

Gyula Varadi, Philippe Lory, David Schultz, Maria Varadi & Arnold Schwartz*

Department of Pharmacology and Cell Biophysics, University of Cincinnati, 231 Bethesda Avenue, Cincinnati, Ohio 45267-0575, USA

THE L-type voltage-dependent calcium channel is an important link in excitation–contraction coupling of muscle cells¹ (reviewed in refs 2 and 3). The channel has two functional characteristics: calcium permeation and receptor sites for calcium antagonists. In skeletal muscle the channel is a complex of five subunits, α_1 , α_2 , β , γ and δ (ref. 4). Complementary DNAs to these subunits have been cloned and their amino-acid sequences deduced^{5–8}. The skeletal muscle α_1 subunit cDNA expressed in L cells manifests as specific calcium-ion permeation, as well as sensitivity to the three classes of organic calcium-channel blockers^{9,10}. We report here that coexpression of the α_1 subunit with other subunits results in significant changes in dihydropyridine binding and gating properties. The available number of drug receptor sites increases 10-fold with an $\alpha_1\beta$ combination, whereas the affinity of the dihydropyridine binding site remains unchanged. Also, the presence of the β subunit accelerates activation and inactivation kinetics of the calcium-channel current.

LCa.11 cells⁹ were cotransfected with various combinations of expression plasmids for α_2 , β and γ subunit cDNAs. Northern blot analyses of poly(A)⁺ RNA preparations from the various transfected cell lines (Fig. 1a, b) confirmed that the parental Ltk⁺ cells expressed no transcripts for the skeletal muscle L-type calcium-channel subunits. The LCa.11 cell line and cell lines derived from LCa.11 retained the ability to express α_1 subunit messenger RNA to a similar extent. Some of the transfectants generated, coexpressed the subunits in equimolar ratios ($L\alpha_1\alpha_2/3$, $L\alpha_1\alpha_2/23$, $L\alpha_1\gamma/6$, $L\alpha_1\beta\gamma/1$, $L\alpha_1\alpha_2\beta\gamma/7$), whereas others ($L\alpha_1\beta/6$, $L\alpha_1\gamma/8$) overexpressed the β and γ subunits (Fig. 1a, b).

Membranes from transfected cells were assayed for binding of [³H](+)-PN200-110, a high-affinity dihydropyridine (DHP) ligand (Table 1). LCa.11 cell membranes showed saturation binding with K_d and B_{max} values of 0.38 ± 0.05 nM and 75 ± 10 fmol mg⁻¹ protein, respectively. The cell lines expressing various combinations of subunits showed a similar affinity to the DHP ligand (Table 1). Although a 50–70% variation among the K_d values for selected combinations of subunit expression was noted, these are not significant as K_d values typically vary from 0.1–0.6 nM, for skeletal muscle membranes¹¹. The number of DHP receptor sites changed when the α_1 subunit was expressed in combination with other subunits. For example, the B_{max} for $L\alpha_1\alpha_2/3$ cell membranes was 112 fmol mg⁻¹ protein compared with 75 fmol mg⁻¹ protein for LCa.11. There was a large increase in the case of the $L\alpha_1\beta/6$ membranes: the B_{max}

* To whom correspondence should be addressed.

TABLE 1 Comparison of DHP-receptor function and calcium-channel activity of L cells expressing combinations of calcium-channel subunits

Cell line	Subunit composition	K_d (nM) ±s.e.m.	B_{max} (fmol mg ⁻¹ protein) ±s.e.m.	I peak (pA)	Increase after 1 μ M Bay K 8644 (%)
LCa.11	α_1	0.38 ± 0.05	75 ± 10 (5)	30.8 ± 7.1 (17)	104 ± 21
L $\alpha_1\alpha_2$ /3	$\alpha_1\alpha_2$	0.41*	112*	37.0 ± 6.8 (10)	99 ± 31
L $\alpha_1\beta$ /6	$\alpha_1\beta$	0.48 ± 0.09	712 ± 108 (4)	14.5 ± 1.2 (23)	-3 ± 7
L $\alpha_1\gamma$ /6	$\alpha_1\gamma$	0.42 ± 0.11	57 ± 3 (3)	11.0 ± 1.5 (7)	49 ± 35
L $\alpha_1\gamma$ /8	$\alpha_1\gamma$	0.51*	44*	—	—
L $\alpha_1\beta\gamma$ /1	$\alpha_1\beta\gamma$	0.57*	39*	22.0 ± 3.1 (14)	6 ± 5
L $\alpha_1\alpha_2\beta$ /10	$\alpha_1\alpha_2\beta$	—	—	20.0 ± 3.0 (18)	36 ± 9
L $\alpha_1\alpha_2\beta\gamma$ /7	$\alpha_1\alpha_2\beta\gamma$	0.64 ± 0.03	192 ± 36 (3)	26.9*	62*

Cell lines are indicated by subunits present and clone number. Saturation binding was performed on crude membrane preparations (protein concentration 50–150 μ g ml⁻¹) in 1-ml assays using the dihydropyridine [³H]-PN200–110 (NEN, USA) in concentrations from 0.1 to 1.5 nM under conditions previously described¹⁰. K_d and B_{max} values were calculated using the nonlinear curve-fitting program LIGAND²³. The Ba²⁺ currents are recorded as described in the legend to Fig. 2. Peak Ba²⁺ currents were measured before and after application of Bay K 8644. Control Ba²⁺ currents and percentage increase induced by Bay K 8644 are given as mean ± s.e.m. Numbers in parenthesis indicate the independent determinations both for drug binding and for electrophysiological studies.

* Mean value of two determinations.

was 10-fold higher (712 ± 108 fmol mg⁻¹ protein) than that in LCa.11 cells. Coexpression of α_1 and γ -specific mRNAs in two cell lines, differing in the relative expression of the γ subunit, caused no change either in affinity for the ligand or in B_{max} . Additionally, no significant difference was seen in B_{max} , comparing an $\alpha_1\beta\gamma$ clone with the LCa.11 cell line. When four different subunits (L $\alpha_1\alpha_2\beta\gamma$ /7) were coexpressed a significant increase in B_{max} was noted, but to a lesser extent than that observed with $\alpha_1\beta$ (L $\alpha_1\beta$ /6).

Taken together, these experiments emphasize the contribution of the β subunit to the DHP receptor function of the α_1 subunit. Thus, it is likely that the β subunit increases the number of functionally active binding sites. This effect may be due to an

enhancement of membrane insertion and/or to stabilization of a specific conformation of the α_1 subunit in the membrane.

L cells transfected with α_1 (Fig. 2a) or with a combination of $\alpha_1\alpha_2$ -expressing plasmids displayed voltage-dependent calcium-channel activity (Table 1). L-type calcium-channel activity was easily detectable in the cells, on the basis of the sensitivity to the DHP agonist, Bay K 8644, and to blockade by 1 mM Cd²⁺ ions (data not shown). Application of a depolarizing pulse to +30 mV, elicited a slowly activating Ba²⁺ current. The complete current-voltage (I/V) curve analysis revealed that this current was activated near +10 mV and peaked at +30 mV (Fig. 2e). In $\alpha_1\alpha_2$ -expressing cells we observed L-type calcium-channel activity of very similar characteristics, with the exception that

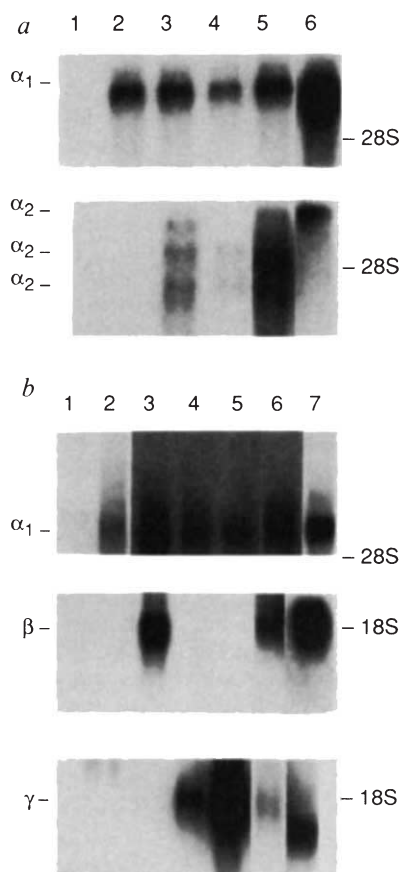


FIG. 1 Coexpression of skeletal muscle L-type calcium-channel subunits in transfected Ltk⁻ cells. Poly(A)⁺ RNA samples²⁴ (4 μ g for each lane) from different cell lines were electrophoresed on 1% formaldehyde-agarose gels²⁵, blotted onto nitrocellulose membrane and hybridized with subunit-specific probes. a, Lane 1, Ltk⁻; lane 2, LCa.11; lane 3, L $\alpha_1\alpha_2$ /3; lane 4, L $\alpha_1\alpha_2$ /21; lane 5, L $\alpha_1\alpha_2$ /23; lane 6, rabbit skeletal muscle. The blot was hybridized either with a labelled full-length α_1 or α_2 probe⁶. b, Lane 1, Ltk⁻; lane 2, LCa.11; lane 3, L $\alpha_1\beta$ /6; lane 4, L $\alpha_1\gamma$ /6; lane 5, L $\alpha_1\gamma$ /8; lane 6, L $\alpha_1\beta\gamma$ /1; lane 7, rabbit skeletal muscle. The blots were hybridized with full-length labelled α_1 , β and γ probes, respectively. Migration of 18S and 28S rRNA markers is shown.

METHODS. LCa.11 cells were transfected with 20 μ g pMAMneo α 2D5'D3' or cotransfected²⁶ with 2 μ g pSV2neo (ref. 27) and 20 μ g pSG-5 β , pSkCaChG3 (ref. 8), respectively. In some cases the equimolar combination of β and γ expression plasmids were used for transfections. Clones surviving in HAT and G-418 media were tested for coexpression of skeletal muscle α_1 , α_2 , β and γ subunit-specific mRNAs. The expression plasmids were constructed as follows. The pMAMneo α 2D5'D3' was assembled from the SmCaCha2.15 and SmCaCha2.3 components of the α_2 cDNA. A 540-base pair (bp) *Sma*I–*Bgl*II and a 5.2-kilobase (kb) *Bgl*II–*Sph*I fragment were isolated from SmCaCha2.15 and SmCaCha2.3, respectively, and ligated into the *Sma*I–*Sph*I sites of pBI24. The resultant 5.7-kb cDNA was cleaved with *Eco*RI and the 3.7-kb-long α_2 coding frame filled in and ligated into the *Sma*I site of the pMAMneo vector (Clontech). The pSG-5 β , an expression plasmid for the β subunit, was constructed from two overlapping cDNAs for the β subunit pSkCaCh β D and pSkCaCh β B, respectively. The cDNAs were isolated from a skeletal muscle-specific Okayama–Berg cDNA library²⁸. The cDNAs, consisting of an open reading frame of 1,572 nucleotides, are identical to that published previously⁷. The pSkCaCh β D starts from the 5'-end of the coding region to nucleotide 1,296. The pSkCaCh β B starts at nucleotide 26 and extends 3' to the poly(A) tail. The pSG-5 β expression plasmid was constructed as follows. A 911-bp *Not*I–*Hind*III fragment and a 0.9-kb *Hind*III–*Hind*III fragment isolated from pSkCaCh β D and pSkCaCh β B plasmids, respectively, were ligated between the *Not*I–*Hind*III sites of pBluescriptKS(+) vector (Stratagene). The full-length β -coding region was liberated by *Not*I and *Pvu*II cleavage, the recessive termini filled in with Klenow polymerase, *Eco*RI linkers added and ligated into the *Eco*RI cloning site of the pSG-5 vector (Stratagene).

the peak currents were slightly higher than that observed with α_1 alone (Table 1). Cells coexpressing only the $\alpha_1\gamma$ subunits had similar L-type calcium channel activity (Fig. 2b). The currents were smaller in amplitude (Table 1) than the α_1 current, but were still enhanced by Bay K 8644.

Unexpectedly and contrary to the DHP binding data, Ba^{2+} currents induced by the coexpression of $\alpha_1\beta$ subunits displayed even smaller amplitudes (Table 1). Moreover, using the $L\alpha_1\beta/6$ clone, modulation of Ba^{2+} current was undetectable with 1 μM Bay K 8644 (Fig. 2d). Despite the lack of a stimulatory effect of Bay K 8644, there was a leftward shift of the I/V curve, compared with α_1 alone (data not shown). These phenomena were fully conserved in a cell line ($L\alpha_1\beta\gamma/1$) coexpressing the $\alpha_1\beta\gamma$ subunits. This cell line (Fig. 2c), which had higher Ba^{2+}

current amplitude, was used to quantitatively study the $\alpha_1\beta$ interaction. The I/V curve, in the absence of Bay K 8644, revealed a leftward shift (10 mV), compared with the α_1 -expressed Ba^{2+} current. The threshold of activation was at 0 mV and the peak current at +20 mV (Fig. 2f). An additional 10 mV leftward shift occurred after the application of 1 μM Bay K 8644, which indicates that the agonist acts on the α_1 subunit¹². But in the presence of the β subunit the dramatic effect on current amplitude enhancement could not be detected (Fig. 2c, d, f). Taken together, these results emphasize the regulatory function of the β subunit on voltage-dependent calcium-channel activity. The inability of Bay K 8644 to enhance the peak current activity clearly appears to be $\alpha_1\beta$ related, particularly when the stoichiometry of coexpression strongly favours β . In addition, the currents expressed by the $\alpha_1\beta$ combination of subunits ran down faster than those with the α_1 subunit alone (Fig. 2c).

The most striking effect of $\alpha_1\beta$ subunit coexpression occurred on activation and inactivation kinetics (Fig. 2c, d). As shown in Fig. 3a, the α_1 -expressed Ba^{2+} current activated at a slower rate than that for skeletal muscle cells^{13,14} (time to peak, 6 s; τ of inactivation, 10 s; $n = 3$); it showed extremely slow inactivation, requiring 60-s pulses for complete inactivation (Fig. 3a). Using a 1.5-s pulse we were unable to observe any inactivation (Fig. 3b). The $\alpha_1\beta\gamma$ subunit-related current displayed a faster activation process (time to peak, 380 ± 39 ms; τ of inactivation, 265 ± 30 ms; $n = 18$), compared with α_1 alone (Fig. 3c). These results strongly implicate the participation of the β subunit in the gating process of the L-type calcium channel.

Our data show that the interaction of homologous, tissue-specific subunits expressed in mammalian cells can cause fundamental changes in calcium-channel kinetics. This effect on kinetics was not observed on heterologous coexpression of channel subunits in *Xenopus* oocytes¹⁵⁻¹⁷. In expression of

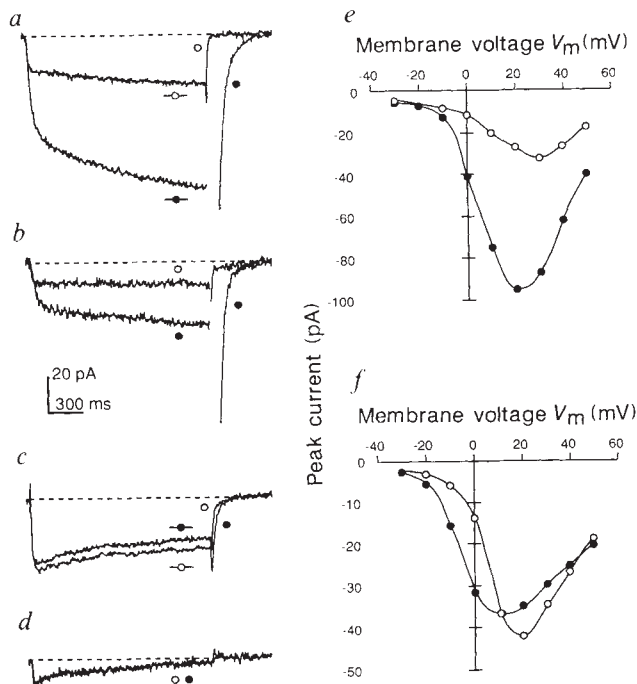


FIG. 2 Electrophysiological and pharmacological characteristics of Ba^{2+} currents elicited on L cells transfected with different combinations of Ca^{2+} channel-subunits. Ba^{2+} currents were recorded on L-cell transfectants using a +30 mV test pulse from a holding potential of -60 mV (a-d), in the absence (○) and in the presence (●) of 1 μM Bay K 8644 (a-d). Typical recordings are presented for: a, LCa.11 cell (α_1 alone); b, $L\alpha_1\gamma/6$ cell (α_1 and γ subunits); c, $L\alpha_1\beta\gamma/1$ cell (α_1 , β and γ subunits); d, $L\alpha_1\beta/6$ (α_1 , β subunits). Corresponding current-voltage relationships are presented in e and f for the LCa.11 cell and the $L\alpha_1\beta\gamma/1$ cell, respectively. The membrane capacitance was 122 pF and 171 pF for the LCa.11 cell and the $L\alpha_1\beta\gamma/1$ cell, respectively.

METHODS. Zn^{2+} induction was performed the day before the measurement of Ca^{2+} channel activity, using a 3-h, 50 μM Zn^{2+} incubation. Fresh culture medium was provided overnight. Just before the electrophysiological experiments, cells were washed in a Tyrode solution (in mM): NaCl 135; CaCl_2 , 1.5; MgSO_4 , 0.4; MgCl_2 , 0.5; KCl, 2; HEPES, 10 (pH adjusted to 7.4 with NaOH). Ionic currents were recorded in the whole-cell configuration²⁹ using an Axopatch 1B amplifier (Axon, California). Ca^{2+} -channel activity was recorded using Ba^{2+} as charge carrier¹². The bath solution contained (in mM): $\text{Ba}(\text{OH})_2$, 40; glutamate, 40; *N*-methyl *D*-glucamine, 80; HEPES, 10; MgCl_2 , 2 (pH adjusted to 7.4 with $\text{CH}_3\text{SO}_3\text{H}$). Pipette solution was (in mM): *N*-methyl *D*-glucamine, 110; MgCl_2 , 2; EGTA, 20; HEPES, 15 (pH adjusted to 7.25 with $\text{CH}_3\text{SO}_3\text{H}$). Ba^{2+} currents were filtered at 1 kHz using a four-pole Bessel filter. Stimulation of the cell, acquisition and analysis of the data were performed using the 5.5 pCLAMP software (Axon, California). Linear currents were subtracted from the Ba^{2+} currents using a P/4 subtraction. Bay K 8644 (racemic mixture) was added directly to the bath solution during the experiments to a final concentration of 1 μM . The application of Cd^{2+} (1 mM) was performed at the end of the experiments to identify precisely the expressed Ca^{2+} -channel activity. No outward currents were observed after this application.

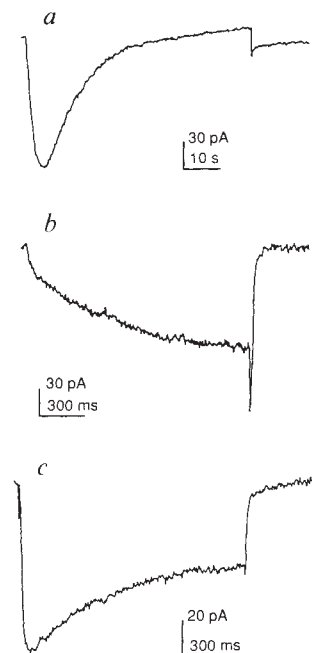


FIG. 3 Kinetic properties of the α_1 and the $\alpha_1\beta\gamma$ Ca^{2+} -channel currents. a, Slow Ba^{2+} current recorded on an LCa.11 cell (membrane capacitance = 395 pF) using a 1 min test pulse (test pulse was +30 mV from a -60 mV holding potential). The P/4 subtraction procedure was not used for this recording. A similar 1.5-s test pulse was used to compare the Ba^{2+} current kinetics on an LCa.11 cell and on an $L\alpha_1\beta\gamma/1$ cell (traces b and c, respectively). A slow activating current was recorded on the LCa.11 cell (trace b; $\tau > 2$ s; membrane capacitance = 340 pF). Ba^{2+} current recorded on the $L\alpha_1\beta\gamma/1$ cell (membrane capacitance = 187 pF) displayed inactivation (c). The peak current was obtained in 200 ms, and the time constant of inactivation was $\tau = 560$ ms.

heterologous subunits, it is likely that the reported effects are due to stabilization of the protein complex in the membrane. Consistent with this notion are the findings that unique tissue-specific isoforms exist for the α_1 (refs 5, 15–20), as well as for the β and γ subunits^{7,8,21}. Hence, these other subunits probably interact with critical regions of the α_1 subunit. Our data show that, of the subunits studied, β appears to be most important in regulating calcium-channel function and the density of the DHP binding region, behaviour interestingly reminiscent of skeletal muscle, in that the high density of DHP-binding sites does not reflect the calcium-channel activity²². Homologous expression strategies and subunits with specifically engineered molecular characteristics should yield a framework for calcium-channel function. □

Received 25 April; accepted 31 May 1991.

1. Rios, E. & Brum, G. *Nature* **325**, 717–720 (1987).
2. Hess, P. A. *Rev. Neurosci.* **13**, 337–356 (1990).
3. Tsien, R. W. & Tsien, R. Y. *An. Rev. Cell Biol.* **6**, 715–760 (1990).
4. Vaghy, P. L., McKenna, E., Itagaki, K. & Schwartz, A. *Trends pharmacol. Sci.* **9**, 398–402 (1988).
5. Tanabe, T. *et al. Nature* **328**, 313–318 (1987).
6. Ellis, S. B. *et al. Science* **241**, 1661–1664 (1988).
7. Ruth, P. *et al. Science* **245**, 1115–1118 (1989).
8. Jay, S. D. *et al. Science* **248**, 490–492 (1990).
9. Perez-Reyes, E. *et al. Nature* **340**, 233–236 (1989).
10. Kim, H. S. *et al. J. Biol. Chem.* **265**, 11858–11863 (1990).
11. Janis, R. A., Silver, P. J. & Triggie, D. J. *Adv. Drug Res.* **16**, 309–591 (1987).
12. Lory, P., Rassendren, F. A., Richard, S., Tiaho, F. & Nargeot, J. *J. Physiol., Lond.* **429**, 95–112 (1990).
13. Caffrey, J. D., Brown, A. M. & Schneider, M. D. *Science* **236**, 570–573 (1987).
14. Caffrey, J. D., Brown, A. M. & Schneider, M. D. *J. Neurosci.* **9**, 3443–3453 (1989).
15. Mikami, A. *et al. Nature* **340**, 230–233 (1989).
16. Biel, M. *et al. FEBS Lett.* **269**, 409–412 (1990).
17. Mori, Y. *et al. Nature* **350**, 398–402 (1991).
18. Sligh, D. F. *et al. FEBS Lett.* **250**, 509–514 (1989).
19. Koch, W. J., Ellinor, P. T. & Schwartz, A. *J. Biol. Chem.* **265**, 17786–17791 (1990).
20. Perez-Reyes, E., Wei, X., Castellano, A. & Birnbaumer, L. *J. Biol. Chem.* **265**, 20430–20436 (1990).
21. Pragnell, M., Leveille, C. F., Jay, S. D. & Campbell, K. P. *Biophys. J.* **59**, 390 (1991).
22. Schwartz, L. M., McCleskey, E. W. & Almers, W. *Nature* **314**, 747–751 (1985).
23. McPherson, G. A. *KINETIC, EBD, LIGAND, LOWRY* (Elsevier, UK, 1985).
24. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
25. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory, New York, 1989).
26. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).
27. Lee, F., Mulligan, R., Berg, P. & Ringold, G. *Nature* **294**, 228–232 (1981).
28. Okayama, H. & Berg, P. *Molec. cell. Biol.* **3**, 280–289 (1983).
29. Hamill, O. P. *et al. Pflügers Arch.* **391**, 85–100 (1981).

ACKNOWLEDGEMENTS. We thank L. Birnbaumer (Baylor College of Medicine) and K. P. Campbell (University of Iowa) for the LCa_v1.1 cells and the skeletal muscle β and γ cDNAs, respectively, E. T. Wallick (University of Cincinnati) for his contributions in the binding studies, and S. Tang (University of Cincinnati) for the preliminary binding experiments. This work was supported by the NIH and Marion Merrell Dow Company (A.S.). P.L. was supported by a Postdoctoral Fellowship from INSERM, France.

Activation of Ca²⁺ entry into acinar cells by a non-phosphorylatable inositol trisphosphate

G. St J. Bird, M. F. Rossier, A. R. Hughes, S. B. Shears, D. L. Armstrong & J. W. Putney Jr

Calcium Regulation Section, Laboratory of Cellular and Molecular Pharmacology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA

IN many cell types, receptor activation of phosphoinositidase C results in an initial release of intracellular Ca²⁺ stores followed by sustained Ca²⁺ entry across the plasma membrane. Inositol 1,4,5-trisphosphate is the mediator of the initial Ca²⁺ release¹, although its role in the mechanism underlying Ca²⁺ entry remains controversial^{2–6}. We have now used two techniques to introduce inositol phosphates into mouse lacrimal acinar cells and measure their effects on Ca²⁺ entry: microinjection into cells loaded with Fura-2, a fluorescent dye which allows the measurement of intra-

cellular free calcium concentration by microspectrofluorimetry, and perfusion of patch clamp pipettes in the whole-cell configuration while monitoring the activity of Ca²⁺-activated K⁺ channels as an indicator of intracellular Ca²⁺. We report here that inositol 1,4,5-trisphosphate serves as a signal that is both necessary and sufficient for receptor activation of Ca²⁺ entry across the plasma membrane in these cells.

Stimulation of muscarinic-cholinergic receptors on a lacrimal acinar cell (Fig. 1a) results in a biphasic increase in intracellular calcium ion concentration ([Ca²⁺]_i) consisting of both an intracellular Ca²⁺ release, followed by the entry of Ca²⁺ across the plasma membrane. Microinjection of heparin, an antagonist of the inositol trisphosphate receptor⁷, blocked both components of the methacholine-induced Ca²⁺ response (Fig. 1b), but did not block the actions of thapsigargin (Fig. 1b). Thapsigargin releases intracellular Ca²⁺ and activates Ca²⁺ entry across the plasma membrane through the same pathway as phospholipase C-linked agonists, but bypasses the production of inositol phosphates^{8,9}. The failure of heparin to inhibit the response to thapsigargin demonstrates that although the actions of inositol 1,4,5-trisphosphate (Ins(1,4,5)P₃) are blocked, the fundamental machinery for Ca²⁺ entry remains intact. Thus, the interaction of Ins(1,4,5)P₃ with one or more intracellular receptors seems to be necessary for activation of sustained Ca²⁺ entry by phosphoinositidase C-linked agonists.

But electrophysiological studies^{10–12} using the activity of Ca²⁺-activated K⁺ channels as an indicator of [Ca²⁺]_i, and internal application of inositol phosphates by perfusion of patch pipettes in whole-cell configuration, were unable to obtain sustained responses to Ins(1,4,5)P₃ alone. Although Ins(1,4,5)P₃ was necessary for a sustained Ca²⁺ signal, it alone was not sufficient; inositol 1,3,4,5-tetrakisphosphate (Ins(1,3,4,5)P₄), a metabolite of Ins(1,4,5)P₃, was also required^{10–12}. By contrast, when we perfused patch pipettes with higher concentrations of Ins(1,4,5)P₃ than those previously used, we obtained sustained activation of the potassium current (Fig. 2c) which was entirely dependent on the presence of extracellular Ca²⁺. Other studies have also shown that Ins(1,4,5)P₃ alone stimulates Ca²⁺ entry in this and other cell types^{6,13–15}.

One difficulty in interpreting experiments with Ins(1,4,5)P₃ is that most cell types, including lacrimal cells, have substantial Ins(1,4,5)P₃ 3-kinase activity⁹. Thus, in the experiment shown in Fig. 2c, some of the internally perfused Ins(1,4,5)P₃ may have been phosphorylated to Ins(1,3,4,5)P₄. Furthermore, when investigating the interactions of Ins(1,4,5)P₃ and Ins(1,3,4,5)P₄ in the physiological concentration range, experimental results could be complicated by the well-known ability of Ins(1,3,4,5)P₄ to inhibit the degradation of Ins(1,4,5)P₃. Therefore, we investigated the actions of inositol 2,4,5-trisphosphate (Ins(2,4,5)P₃). This poorly metabolized^{16–19} analogue of Ins(1,4,5)P₃ releases the same intracellular Ca²⁺ pool as Ins(1,4,5)P₃ (refs 20, 21), but there are conflicting reports about its ability to activate Ca²⁺ entry^{6,11,13,17}. Before testing the effects of Ins(2,4,5)P₃ on Ca²⁺ entry, we examined the ability of mouse lacrimal cell Ins(1,4,5)P₃ 3-kinase to phosphorylate Ins(2,4,5)P₃. The lacrimal enzyme rapidly phosphorylated ³H-labelled Ins(1,4,5)P₃ but no phosphorylation of ³H-labelled Ins(2,4,5)P₃ was detectable (Fig. 3). Thus, we conclude that phosphorylation of Ins(2,4,5)P₃ by lacrimal cell Ins(1,4,5)P₃ 3-kinase proceeds slowly if at all, and it is unlikely that internally applied Ins(2,4,5)P₃ will result in the accumulation of significant quantities of a 3-phosphorylated InsP₄.

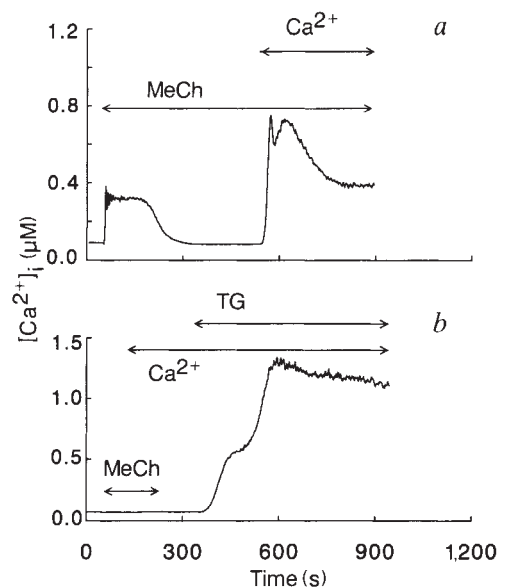
When patch pipettes were perfused with Ins(2,4,5)P₃ there was a sustained activation of potassium conductance (Fig. 2d). This effect of Ins(2,4,5)P₃ was inhibited reversibly by removal of extracellular Ca²⁺, and was blocked by perfusion of the pipette with 1 mg ml⁻¹ heparin (not shown). Because sustained Ca²⁺ signalling was activated by Ins(2,4,5)P₃, which is not phosphorylated, it seems that the interaction of InsP₃ with its intracellular receptor is sufficient for the activation of Ca²⁺ entry

without formation of $\text{Ins}(1,3,4,5)\text{P}_4$. This conclusion is difficult to reconcile with studies using the same mouse lacrimal acinar cell system and the perfused patch-pipette technique that indicated that $\text{Ins}(1,3,4,5)\text{P}_4$ is required for a sustained $[\text{Ca}^{2+}]_i$ signal¹⁰⁻¹². In the whole-cell configuration, soluble factors will be lost by dialysis from the cytoplasm, and it is possible that

these factors might be depleted to different extents in different laboratories. In the previous experiments¹⁰⁻¹², substances could have been lost whose function can be supplanted by $\text{Ins}(1,3,4,5)\text{P}_4$, or in our experiments factors necessary for the action of $\text{Ins}(1,3,4,5)\text{P}_4$ could have been lost. Thus, we next directly microinjected $\text{Ins}(2,4,5)\text{P}_3$ into intact lacrimal cells, with

FIG. 1 Calcium signalling in single, microinjected mouse lacrimal acinar cells. *a*, An acinar cell, previously injected with Fura-2, was activated in the absence of extracellular Ca^{2+} by the addition of $1\ \mu\text{M}$ methacholine (MeCh). Where indicated, extracellular Ca^{2+} was restored to $1.8\ \text{mM}$. The biphasic pattern of $[\text{Ca}^{2+}]_i$ signalling confirms previous observations on populations of lacrimal acinar cells^{9,22}. *b*, A single cell, previously injected with a solution containing $2\ \text{mM}$ Fura-2 and $100\ \text{mg}\ \text{ml}^{-1}$ low molecular weight heparin (Sigma), was activated with methacholine under the same experimental conditions as in *a*. No response was seen on addition of methacholine and on addition of Ca^{2+} , but a normal, sustained response was obtained on addition of thapsigargin (TG).

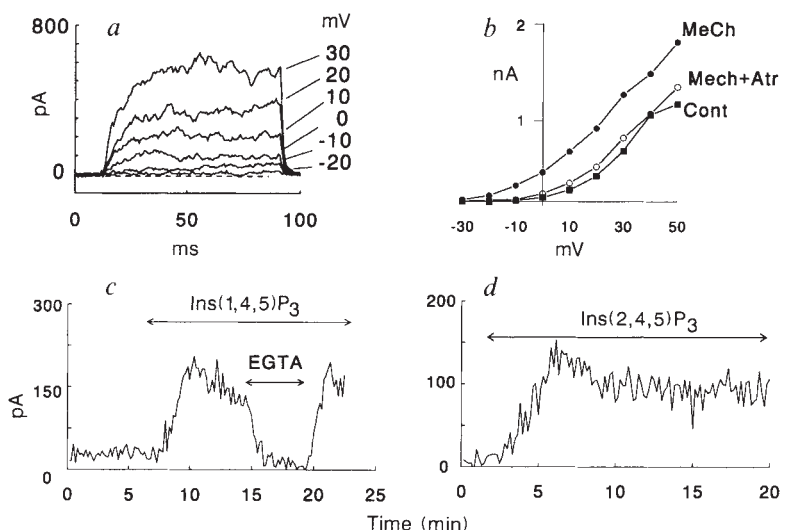
METHODS. Mouse lacrimal acinar cells were prepared essentially as described previously^{9,23}. After isolation, the acinar cells were washed and resuspended in sterile DMEM medium containing 10% fetal bovine serum, $5\ \text{mM}$ glutamine, $50\ \text{U}\ \text{ml}^{-1}$ penicillin and $50\ \text{U}\ \text{ml}^{-1}$ streptomycin. The cells were then allowed to attach to glass coverslips coated with Cell-Tak ($3.5\text{--}10\ \mu\text{g}\ \text{cm}^{-2}$, Biopolymers). Acinar cells were incubated on the glass coverslips for at least 3 h before use. The attached cells were mounted in a Teflon chamber (Bionique, NY) and incubated with $0.5\ \mu\text{M}$ of the acetoxymethyl ester of Fura-2 (Fura-2/AM, Molecular Probes) for 30 min at room temperature. The cells were washed and incubated in a HEPES-buffered physiological saline solution⁹ at room temperature for at least 30 min before $[\text{Ca}^{2+}]_i$ measurements were made. The fluorescence of the Fura-2-loaded cells was monitored with a photomultiplier-based system as previously described²⁴. All experiments were done at room temperature. $[\text{Ca}^{2+}]_i$ values were calculated by using Ca^{2+} -saturated and Ca^{2+} -depleted Fura-2 solutions for R_{max} and R_{min} , respectively, as previously described²⁵. A solution consisting of $27\ \text{mM}\ \text{K}_2\text{HPO}_4$, $8\ \text{mM}\ \text{Na}_2\text{HPO}_4$, $26\ \text{mM}\ \text{KH}_2\text{PO}_4$, pH 7.2 and $2\ \text{mM}$ Fura-2 (acid), was pressure injected into cells, using a glass micropipette attached to a WPI PV830 Picopump (World Precision Instruments, New Haven, Connecticut). In some experiments, this solution also contained either $\text{Ins}(1,4,5)\text{P}_3$ ($10\ \text{mM}$), $\text{Ins}(2,4,5)\text{P}_3$ ($10\ \text{mM}$) (see Fig. 4), or heparin ($100\ \text{mg}\ \text{ml}^{-1}$). Lacrimal cells previously loaded with Fura-2 by incubation with Fura-2/AM were used so that $[\text{Ca}^{2+}]_i$ levels could be monitored before micro-



injection. A successful microinjection was indicated by a further increase in cell fluorescence due to the injected Fura-2. Despite the increase in fluorescence seen, no increase in apparent $[\text{Ca}^{2+}]_i$ was observed in cells microinjected with solutions lacking inositol phosphates. The precise volume of the injections was not determined but the dilution of the injected material was estimated to be of the order of at least 100- and probably 1,000-fold, on the basis of the increase in Fura-2 fluorescence. For these, and all experiments in this report, results of a single experiment are presented which are representative of at least three independent determinations.

FIG. 2 Activation of sustained Ca^{2+} entry by inositol trisphosphates, as assessed by Ca^{2+} -dependent potassium current. Mouse lacrimal acinar cells were voltage-clamped with patch pipettes in the whole-cell configuration. *a*, Leak-subtracted currents elicited by voltage steps from $-50\ \text{mV}$ to the voltages indicated. *b*, Methacholine (MeCh; $10\ \mu\text{M}$) increases the control outward current (Cont) at all voltages. This effect is antagonized by $10\ \mu\text{M}$ atropine (Atr). *c*, Perfusion of the patch pipette with $500\ \mu\text{M}$ $\text{Ins}(1,4,5)\text{P}_3$ produces a sustained increase in outward current measured at $0\ \text{mV}$, which depends completely on extracellular Ca^{2+} . Adding EGTA to the bath reversibly blocks the current. *d*, Perfusion of another cell with $100\ \mu\text{M}$ $\text{Ins}(2,4,5)\text{P}_3$ also produces a sustained increase outward current. This response is also blocked by removal of extracellular Ca^{2+} (not shown).

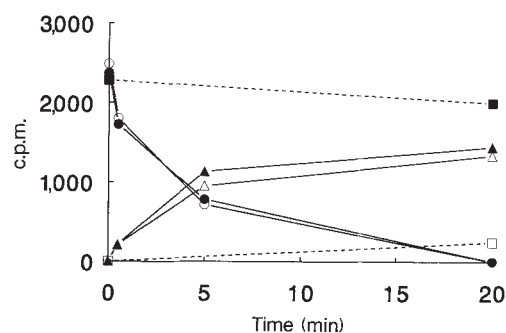
METHODS. The macroscopic current through Ca^{2+} -activated K^+ channels was recorded at room temperature under voltage-clamp with the whole-cell configuration of the patch clamp technique²⁶. The bath solution contained $140\ \text{mM}$ NaCl, $4.7\ \text{mM}$ KCl, $1.13\ \text{mM}$ MgCl_2 , $1.2\ \text{mM}$ CaCl_2 , $40\ \text{mM}$ glucose and $10\ \text{mM}$ HEPES, pH 7.2. The patch pipette ($2\text{--}4\ \text{M}\Omega$, Corning glass, 7052) contained $140\ \text{mM}$ KCl, $4\ \text{mM}$ MgCl_2 , $5\ \text{mM}$ Na_2ATP and $10\ \text{mM}$ HEPES, pH 7.2. The pipette solution also contained 0.5 or $1.0\ \text{mM}$ $\text{K}_4\text{--}1$, 2-Brs (*O*-aminophenoxy)ethane-*N*, *N*, *N'*, *N'*-tetraacetic acid potassium-BAPTA to stabilize the low access resistance ($5\text{--}15\ \text{M}\Omega$) from pipette to cell. The patch pipette was perfused through a small quartz capillary which allowed the solution in direct contact with the inside of the cell to be changed during experiments²⁷. Small cells (membrane capacitance $\sim 15\ \text{pF}$) were voltage-clamped to $-50\ \text{mV}$, and every 5 s the voltage was stepped to $0\ \text{mV}$ for 80 ms. At $0\ \text{mV}$, the net current through other Ca^{2+} -activated channels in lacrimal cells is close to zero under these ionic conditions. After subtraction of leak and capacitive currents by a $-P/4$ method (Pclamp 5.5, Axon Instruments, Burlingame, California), the



resulting current was filtered at $1,000\ \text{Hz}$ and digitized at $200\text{-}\mu\text{s}$ intervals. Outward currents were blocked completely by replacing K^+ in the pipette with Cs^+ , or by adding $1\text{--}5\ \text{mM}$ tetraethylammonium to the bath (not shown). The last 20 ms of the steady-state current was averaged and plotted as a function of time. The bath solution could also be changed by perfusion from a gravity feed. The labels in the figures indicate the exact times when new bath or pipette solutions were introduced, without any correction for the dead time required for a new solution to reach the cell. The dead time was $\sim 1.0\ \text{s}$ for the bath and $1\text{--}5\ \text{min}$ for the pipette, depending on capillary placement.

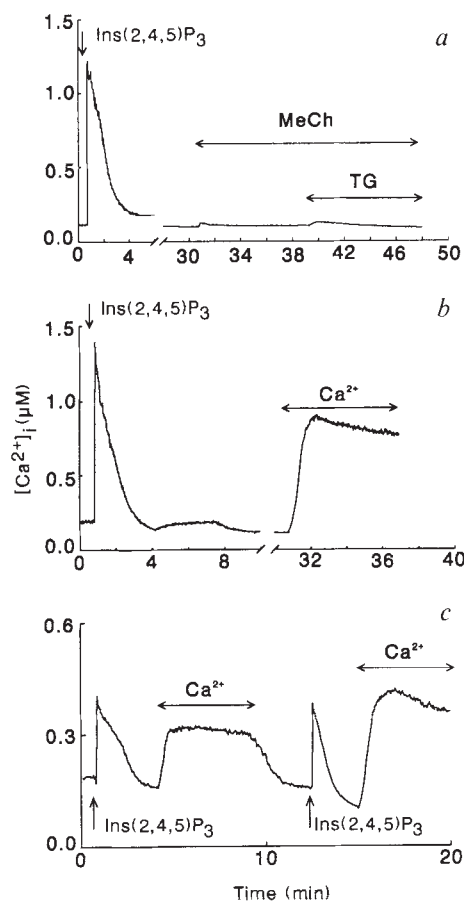
FIG. 3 Metabolism of ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$ and ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ by mouse lacrimal gland supernatants. ■, ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$; □, apparent ^3H -labelled inositol biphosphates derived from ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$. No material running in the InsP_4 region of the chromatogram was detected in samples incubated with ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$. ●, ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$; ▲, ^3H -labelled $\text{Ins}(1,3,4,5)\text{P}_4$ derived from ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$. The difference between accumulated ^3H -labelled $\text{Ins}(1,3,4,5)\text{P}_4$ and disappearance of ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$ is accounted for by small quantities of ^3H -labelled $\text{Ins}(1,3,4)\text{P}_3$ and ^3H -labelled InsP_2 isomers. ○ and △, experiments carried out as for ● and ▲ except that $100\ \mu\text{M}$ ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ was included in the incubations with ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$. Similar rates of metabolism of $\text{Ins}(1,4,5)\text{P}_3$ were obtained in the presence of ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$, thereby insuring that the ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ preparation did not contain any materials which inhibit the 3-kinase. For ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ incubations with lacrimal supernatant (or with whole homogenates, data not shown) no radioactivity was detected in the InsP_4 region of the chromatogram. Phosphorylation of ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ at a rate of 1% of that for ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$ would have been detectable in this experiment. When the supernatant protein concentration was increased 10-fold $\text{Ins}(1,4,5)\text{P}_3$ was converted to $\text{Ins}(1,3,4,5)\text{P}_4$ in seconds but again ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ was not phosphorylated (not shown).

METHODS. Four to ten mice were sacrificed by CO_2 asphyxiation, the lacrimal glands were quickly removed and homogenized (Polytron) on ice in an intracellular-like solution containing 20 mM NaCl, 100 mM KCl, 5 mM Mg_2SO_4 , 20 mM Hepes, 1 mM EGTA, 10 mM phosphocreatine, $10\ \text{U}\ \text{ml}^{-1}$ creatine phosphokinase, 3 mM ATP, $2\ \mu\text{g}\ \text{ml}^{-1}$ leupeptin, 0.2 mM benzamide, $0.2\ \text{mg}\ \text{ml}^{-1}$ phenylmethylsulfonyl fluoride and two calpain inhibitors, E-64 (5 nM) and 5 μM calpain inhibitor I (tripeptide). Enough CaCl_2 was added to give a free-calcium concentration of $1\ \mu\text{M}$. The homogenate was centrifuged for 1 h at $100,000g$ at 2°C and the supernatant removed for analysis of 3-kinase activity. ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$, ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ ($100\ \mu\text{M}$) or a combination of the two was incubated with lacrimal gland supernatant



(30–50 μg protein) in the solution described above for homogenization, for the indicated times. The reactions were stopped with perchloric acid, and the ^3H -labelled materials resolved by HPLC as previously described^{28,29}. ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ was prepared by alkaline hydrolysis of ^3H -labelled phosphatidylinositol 4,5-bisphosphate (NEN)²⁰. The resultant ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ was separated from ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$ and ^3H -labelled inositol 4,5-bisphosphate by HPLC^{28,29} and desalted³⁰. The purity of the final ^3H -labelled material was ascertained to be $\geq 94\%$ by polyol analysis following oxidation with periodate, reduction and dephosphorylation³¹. On other occasions, ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ was prepared by acid-catalysed rearrangement of phosphates by incubating ^3H -labelled $\text{Ins}(1,4,5)\text{P}_3$ in 1 ml aliquots for 60 min at 110°C with 1 M HCl in heat-sealed glass tubes. The resulting preparation was much less pure than that obtained by alkaline hydrolysis of phosphatidylinositol 4,5-bisphosphate (about 56%), but results of experiments with this material were identical to those with the purer material.

FIG. 4 Microinjection of $\text{Ins}(2,4,5)\text{P}_3$ into mouse lacrimal acinar cells. Mouse lacrimal acinar cells previously loaded with Fura-2 by incubation with the acetoxymethyl ester, were further incubated in the absence of external Ca^{2+} and were injected with a solution containing $10\ \text{mM}$ $\text{Ins}(2,4,5)\text{P}_3$ where indicated. This should result in an intracellular $\text{Ins}(2,4,5)\text{P}_3$ concentration of $10\text{--}100\ \mu\text{M}$, which should be sufficient to give maximal or near maximal intracellular Ca^{2+} release (in permeable mouse lacrimal acinar cells, $10\ \mu\text{M}$ $\text{Ins}(2,4,5)\text{P}_3$ is required for half-maximal, and $100\ \mu\text{M}$ for maximal Ca^{2+} release; G.St.J. B., J. Obie and J. W. P. Jr, unpublished observations). $[\text{Ca}^{2+}]_i$ was monitored by microspectrofluorimetry as described for Fig. 1. a, Subsequent addition of $100\ \mu\text{M}$ methacholine (MeCh) followed by $2\ \mu\text{M}$ thapsigargin (TG) failed to induce significant additional Ca^{2+} mobilization. b, Addition of $1.8\ \text{mM}$ Ca^{2+} to the medium 30 min after the injection of $\text{Ins}(2,4,5)\text{P}_3$ induced a substantial sustained elevation of $[\text{Ca}^{2+}]_i$. Similar additions of Ca^{2+} to cells injected with control solution showed no increase in $[\text{Ca}^{2+}]_i$ (not shown). Also, although injection of cells with an equivalent quantity of $\text{Ins}(1,4,5)\text{P}_3$ similarly blocked subsequent intracellular mobilization by methacholine or thapsigargin, this procedure did not result in a sustained response to the addition of Ca^{2+} , presumably due to the rapid intracellular metabolism of the injected $\text{Ins}(1,4,5)\text{P}_3$ (not shown). But similar sustained responses were obtained with other poorly metabolized analogues of $\text{Ins}(1,4,5)\text{P}_3$ ($\text{Ins}(1,4,5)\text{P}_3$), the thio analogue of $\text{Ins}(1,4,5)\text{P}_3$, one experiment, and glycerophosphoinositol 4,5-bisphosphate, three experiments). Note that activation of sustained Ca^{2+} entry by microinjection of $\text{Ins}(2,4,5)\text{P}_3$ could last $>30\ \text{min}$ (Fig. 4b), in one case for 2 h (not shown). This is clearly sufficient time for any inadvertently generated or introduced $\text{Ins}(1,3,4,5)\text{P}_4$ (whose cellular half-life has been estimated to be $\sim <1\ \text{min}$ ²⁹) to be completely metabolized, and also argues against a 'memory' for $\text{Ins}(1,3,4,5)\text{P}_4$ which has been shown to persist for 5–10 min after cessation of $\text{Ins}(1,3,4,5)\text{P}_4$ perfusion in patch clamp experiments^{12,32}. c, $\text{Ins}(2,4,5)\text{P}_3$ was injected where indicated, but the pipette contained only $1\ \text{mM}$ $\text{Ins}(2,4,5)\text{P}_3$ which should give sufficient intracellular $\text{Ins}(2,4,5)\text{P}_3$ to induce submaximal release of Ca^{2+} . Sequential injections of submaximal $\text{Ins}(2,4,5)\text{P}_3$ caused sequential release of Ca^{2+} , and the cumulative effects of these injections resulted in an increased level of Ca^{2+} entry. $\text{Ins}(2,4,5)\text{P}_3$ was obtained from Calbiochem. This material was analysed and repurified by HPLC as follows: about 8,000 d.p.m. each of ^{32}P -labelled $\text{Ins}(1,4,5)\text{P}_3$ and ^3H -labelled $\text{Ins}(2,4,5)\text{P}_3$ were added to $400\ \mu\text{g}$ $\text{Ins}(2,4,5)\text{P}_3$ and the mixture was resolved by HPLC^{28,29}. Fractions containing 84% of ^{32}P (contaminated with 0.3% ^3H) were pooled, as were fractions containing 74% of ^3H (contaminated with 3.8% of the ^{32}P). Each of these pooled fractions was desalted on H^+ -Dowex. Analysis of organic phosphate indicated that the original commercial preparation of $\text{Ins}(2,4,5)\text{P}_3$ was about 94% pure, the balance being primarily



$\text{Ins}(1,4,5)\text{P}_3$. HPLC purification of this material increased the purity to $\geq 99.8\%$. Experiments were done with the commercially available material, as well as the HPLC-purified material, and no qualitative or quantitative differences in experimental results were noted.

the view that such a procedure would not cause the loss of cytoplasmic constituents during the experiment.

Microinjection of $\text{Ins}(2,4,5)\text{P}_3$ in the absence of extracellular Ca^{2+} , caused a rapid and transient release of intracellular Ca^{2+} and depleted the agonist- and thapsigargin-sensitive Ca^{2+} pool (Fig. 4a). Addition of Ca^{2+} to cells previously injected with $\text{Ins}(2,4,5)\text{P}_3$ rapidly increased $[\text{Ca}^{2+}]_i$ to a sustained level, indicating Ca^{2+} entry (Fig. 4b). Injection of a submaximal quantity of $\text{Ins}(2,4,5)\text{P}_3$ resulted in submaximal Ca^{2+} release and a submaximal level of Ca^{2+} entry. Subsequent injection of additional $\text{Ins}(2,4,5)\text{P}_3$ released additional Ca^{2+} and further increased the level of Ca^{2+} entry (Fig. 4c). This indicates that supramaximal activation of the $\text{Ins}(1,4,5)\text{P}_3$ receptor is not required for $\text{Ins}(2,4,5)\text{P}_3$ to activate Ca^{2+} entry.

Our results provide a complete and internally consistent characterization of the role of inositol polyphosphates in receptor-regulated Ca^{2+} entry. Working with mouse lacrimal cells, we have found that (1) the $\text{Ins}(1,4,5)\text{P}_3$ receptor antagonist, heparin, completely blocks both phases of the $[\text{Ca}^{2+}]_i$ signal to methacholine (but not to thapsigargin), (2) intracellular application of $\text{Ins}(2,4,5)\text{P}_3$ fully activates sustained Ca^{2+} entry, and (3) $\text{Ins}(2,4,5)\text{P}_3$ is not phosphorylated by lacrimal cell $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase. Thus we conclude that $\text{Ins}(1,4,5)\text{P}_3$ signals the activation of both phases of $[\text{Ca}^{2+}]_i$ signalling by surface receptor agonists. The mechanism by which $\text{Ins}(1,4,5)\text{P}_3$ activates Ca^{2+} entry presumably results from its ability to deplete intracellular Ca^{2+} stores, as suggested by the capacitative model for Ca^{2+} entry^{2,3}; however, this issue is not specifically addressed by our studies. Although the action of $\text{Ins}(1,3,4,5)\text{P}_4$ observed under certain experimental conditions is an interesting observation, our results with microinjection of $\text{Ins}(2,4,5)\text{P}_3$ into intact cells call into question the physiological relevance of this phenomenon. Under our experimental conditions, and probably

in intact lacrimal cells under physiological conditions, $\text{Ins}(1,4,5)\text{P}_3$ provides a necessary and sufficient signal for both the intracellular Ca^{2+} release and entry of Ca^{2+} across the plasma membrane in receptor-regulated Ca^{2+} signalling. □

Received 27 February; accepted 31 May 1991.

- Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197–205 (1989).
- Putney, J. W. Jr *Cell Calcium* **7**, 1–12 (1986).
- Putney, J. W. Jr *Cell Calcium* **11**, 611–624 (1990).
- Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1990).
- Irvine, R. F. *Biochem. Soc. Trans.* **17**, 6–9 (1989).
- Matthews, G., Neher, E. & Penner, R. J. *Physiol., Lond.* **418**, 105–130 (1989).
- Hill, T. D., Berggren, P.-O. & Boynton, A. L. *Biochem. biophys. Res. Commun.* **149**, 897–901 (1987).
- Thastrup, O. et al. *Agents & Actions* **27**, 17–23 (1989).
- Kwan, C. Y., Takemura, H., Obie, J. F., Thastrup, O. & Putney, J. W. Jr *Am. J. Physiol.* **258**, C1006–C1015 (1990).
- Morris, A. P., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *Nature* **330**, 653–655 (1987).
- Changya, L., Gallacher, D. V., Irvine, R. F., Potter, B. V. L. & Petersen, O. H. *J. Membrane Biol.* **109**, 85–93 (1989).
- Changya, L., Gallacher, D. V., Irvine, R. F. & Petersen, O. H. *FEBS Lett.* **251**, 43–48 (1989).
- Snyder, P. M., Krause, K.-H. & Welsh, M. J. *J. biol. Chem.* **263**, 11048–11051 (1988).
- Yamaguchi, K., Hirata, M. & Kuriyama, H. *Biochem. J.* **251**, 129–134 (1988).
- Llano, I., Marty, A. & Tanguy, J. *Pflügers Arch.* **409**, 499–506 (1987).
- Ryu, S. H., Lee, S. Y., Lee, K.-Y. & Rhee, S. G. *FASEB J.* **1**, 388–393 (1987).
- Irvine, R. F. & Moor, R. M. *Biochem. J.* **240**, 917–920 (1986).
- Morris, A. J., Murray, K. J., England, P. J. & Downes, C. P. *Biochem. J.* **251**, 157–163 (1988).
- Hill, T. D., Dean, N. M. & Boynton, A. L. *Science* **242**, 1176–1178 (1988).
- Irvine, R. F., Brown, K. D. & Berridge, M. J. *Biochem. J.* **221**, 269–272 (1984).
- Spat, A., Bradford, P. G., McKinney, J. S., Rubin, R. P. & Putney, J. W. Jr *Nature* **319**, 514–516 (1986).
- Parod, R. J. & Putney, J. W. Jr *J. Physiol., Lond.* **281**, 371–381 (1978).
- Parod, R. J., Leslie, B. A. & Putney, J. W. Jr *Am. J. Physiol.* **239**, G99–105 (1980).
- Burgess, G. M., Bird, G. St J., Obie, J. F. & Putney, J. W. Jr *J. biol. Chem.* (in the press).
- Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. biol. Chem.* **260**, 3440–3450 (1986).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
- Tang, J. M., Wang, J., Quandt, F. N. & Eisenberg, R. S. *Pflügers Arch.* **416**, 347–350 (1990).
- Dean, N. M. & Moyer, J. D. *Biochem. J.* **242**, 361–366 (1987).
- Hughes, A. R., Takemura, H. & Putney, J. W. Jr *J. biol. Chem.* **263**, 10314–10319 (1988).
- Shears, S. B., Evans, W. H., Kirk, C. J. & Michell, R. H. *Biochem. J.* **256**, 363–369 (1988).
- Stephens, L. R., Hawkins, P. T. & Downes, C. P. *Biochem. J.* **262**, 727–737 (1989).
- Kennedy, E. D., Challis, R. A. J. & Nahorski, S. R. *J. Neurochem.* **53**, 1652–1655 (1989).

ACKNOWLEDGEMENTS. M.F.R. was supported in part by the Fondation Suisse de Bourses en Médecine et Biologie.

v-Src and EJ Ras alleviate repression of c-Jun by a cell-specific inhibitor

Vijay R. Baichwal, Adam Park & Robert Tjian

Howard Hughes Medical Institute & Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA

THE AP-1 family of transcription factors, which includes the proto-oncogene products c-Jun and c-Fos¹, controls the stimulation of cellular genes by growth factors and the expression of oncogenes, including *src* and *ras*^{2–8}. Transcriptional activation by c-Jun is regulated by a cell-type-specific inhibitor that represses the activity of a transcriptional activation domain (A1) of c-Jun by operating through the adjacent negative regulatory region (δ) (refs 9–11). Here we show that cotransfection of the *src* or *ras* oncogene enhances the transcriptional activity of a GAL4:c-Jun hybrid that includes the δ -A1 region of c-Jun, suggesting that the DNA binding and dimerization domain of c-Jun is not required for stimulation by Src or Ras. Moreover, induction of c-Jun activity by Src and Ras occurs in cell lines containing the c-Jun inhibitor but not in a cell line lacking it. The region in c-Jun essential for the stimulatory action of these oncogenes maps to domain A1. These findings suggest the existence of signal-transduction pathways that result in an increase in transcriptional activity of c-Jun and AP-1 by disrupting the c-Jun:inhibitor interaction.

An *in vivo* competition assay, previously used to detect the c-Jun inhibitor in HeLa cells lacking thymidine kinase activity (TK⁻ cells)⁹, was used to screen other cell lines for the presence of this inhibitor. Two murine cell lines, NIH3T3 and L-M (TK⁻),

and primary chicken embryo fibroblasts (CEF) were found to possess a c-Jun inhibitory activity (data not shown). As AP-1 activity is induced by the expression of *src* and *ras* in these murine cell lines⁷, we determined whether this induction is caused by antagonism of inhibition. To ensure we could measure activation by the c-Jun transcriptional activation domains specifically, without potential contribution from changes in other members of the AP-1 family, a GAL4:c-Jun hybrid was constructed⁹ that contains the regulatory region δ and transcriptional activation domains A1 and A2 of c-Jun fused to the DNA-binding domain of the yeast transcriptional activator GAL4. This chimera and a GAL4 reporter plasmid were cotransfected into L-M (TK⁻) cells along with an expression vector carrying the v-*src* or the EJ *ras* gene. Both *src* and *ras* enhanced activation by the GAL4:Jun hybrid (GAL4:Jun 5–253) that contained δ , A1 and A2 domains, but not activation by a similar GAL4:Jun hybrid (GAL4:Jun 139–253) that included the A2 activation domain but lacked the regulated δ -A1 region of Jun (Fig. 1a). The increase in activation by GAL4:Jun 5–253 is not the result of an increase in the amount of this fusion protein, as both chimaeric Jun proteins are expressed at similar levels as measured by a western blot (data not shown). In contrast to the effect in L-M (TK⁻) cells, v-*src* and EJ *ras* had no stimulatory effect in HepG2 cells, a human hepatoma cell line that lacks the c-Jun inhibitory activity⁹ (Fig. 1b). Together these results strongly suggest that *src* and *ras* enhance the transcriptional activity of c-Jun by decreasing the interaction of c-Jun with its inhibitor, thereby relieving repression of the δ -A1 region of c-Jun.

The v-Jun protein, which lacks the δ region, can still be partially regulated by the c-Jun inhibitor⁹. We therefore determined whether v-Src and EJ Ras could stimulate the transcriptional activity of a v-Jun E2 hybrid that has the transcriptional activation domains of v-Jun and the DNA-binding domain of

the E2 transcriptional activator from bovine papilloma virus. Both oncogenes enhanced the transcriptional activity of this hybrid in L-M (TK⁻) cells but not HepG2 cells (Fig. 1c), suggesting that these oncogenes enhanced transcriptional activity at least in part by affecting the interaction of the inhibitor with the A1 domain, even in the absence of δ .

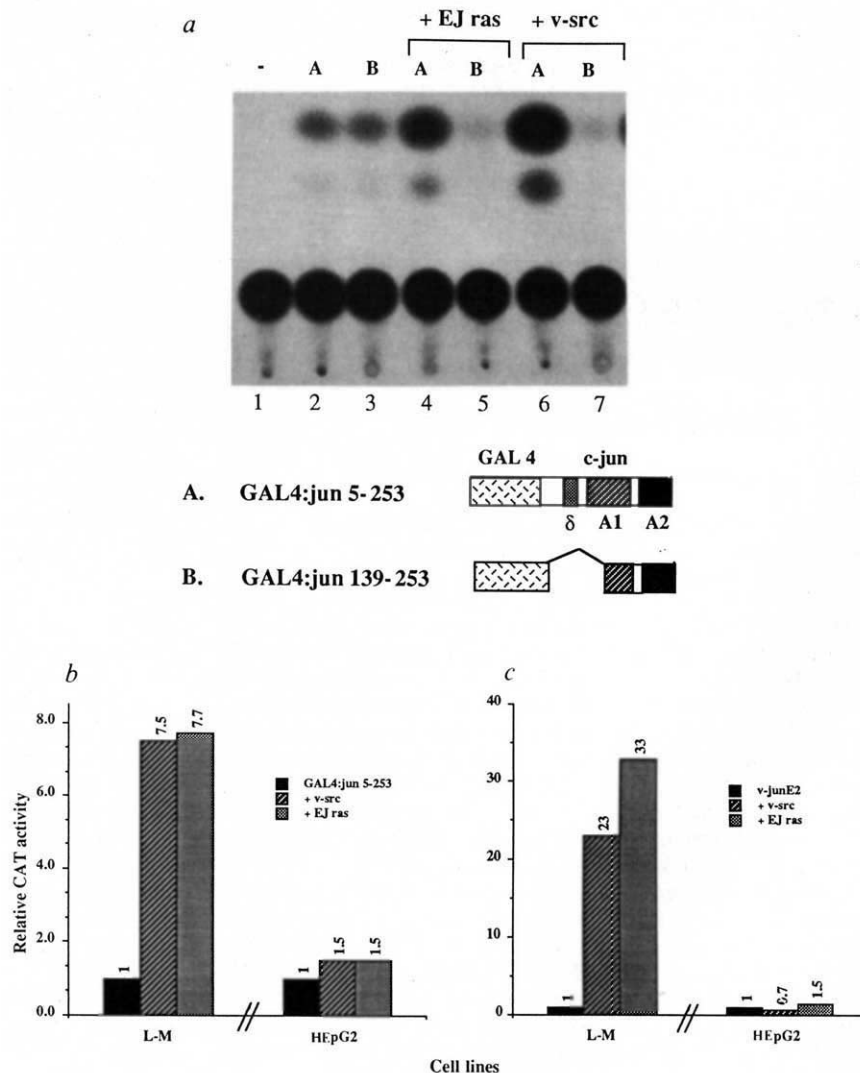
To confirm that the stimulatory effect of v-Src resulted from an increase in activity of the A1 domain in c-Jun, several GAL4:Jun hybrids were examined for stimulation by v-Src in L-M (TK⁻) and in primary CEF, a cell type that can be transformed by v-Src. In the absence of cotransfected Src, all hybrids, including the one that had only a functional A2 domain, activated the reporter plasmid (GAL4:Jun 139–253; 15- and 3.4-fold in L-M (TK⁻) and CEF, respectively; Fig. 2a). Transfection with v-Src, however, significantly enhanced transcriptional activity of only those hybrids that had an intact A1 domain (GAL4:Jun 5–253, GAL4:Jun 69–253 and GAL4:Jun 5–154), but not that of the hybrid with only a functional A2 domain (GAL4:Jun 139–253). Thus in both cell types, the target sequence necessary for Src stimulation mapped to the A1 domain of c-Jun. The identification of a specific target region in the jun protein that is necessary for the stimulatory effect of v-Src also establishes that the v-Src-induced increase in transcriptional activity of c-Jun is not merely caused by some general stimulation of

transcription by v-Src in these cells. Examination of the ability of various v-Src mutants¹² to enhance the transcriptional activity of GAL4:Jun 5–253 showed that induction by v-Src depended on its tyrosine kinase activity (mutant SHX14, Fig. 2b) and proper subcellular location (mutant MM4, Fig. 2b) in both L-M (TK⁻) and CEF. Mutations in the SH2 region (mutants SHX13, SPX1 and XD4) had a host cell-dependent effect on induction by v-Src. This region of Src is involved in the interaction of v-Src with a phosphatidylinositol kinase, the Ras-specific GTPase-activating protein (GAP) and the GAP-associated protein p62 (refs 13, 14). These findings therefore suggest that the phosphatidylinositol kinase and/or GAP/p62 might be involved in the effects of v-Src on Jun transcriptional activity.

The stimulatory effect of EJ Ras on the transcriptional activity of the GAL4:Jun hybrid is also seen in NIH3T3 cells, a cell type that can be transformed by EJ Ras. Analysis of induction of various GAL4:Jun hybrids by Ras revealed that, as with v-Src, the A1 domain of Jun is essential for the stimulatory effect of Ras in NIH3T3 cells (Fig. 3). All the GAL4:Jun hybrids examined activated the reporter gene to a similar extent in the absence of cotransfected *ras* (29-, 10-, 4- and 11-fold). But Ras significantly stimulated the transcriptional activity of only the hybrids that had an intact A1 domain. The identification of the A1 domain as a target region necessary for the stimulatory effect

FIG. 1 a, v-Src and EJ Ras enhance activation by GAL4:Jun hybrids in L-M (TK⁻) cells. L-M (TK⁻) cells were transfected with a GAL4 reporter plasmid alone (lane 1) or together with GAL4:Jun 5–253 (lanes 2, 4, and 6), or with GAL4:Jun 139–253 (lanes 3, 5, and 7) and EJ Ras (lanes 4 and 5), or v-Src (lanes 6 and 7). An autoradiogram of the TLC plate used to monitor chloramphenicol acetyl transferase (CAT) activity is shown. Schematic structures of the GAL4:Jun hybrids transfected are shown and the δ , A1 and A2 domains of Jun are indicated. Note that in these and other Jun hybrids the DNA binding and dimerization domains of Jun are deleted. b, v-Src and EJ Ras enhance activation by Gal4:Jun 5–253 in L-M (TK⁻) but not HepG2 cells. A GAL4 reporter was cotransfected with GAL4:Jun 5–253 in the presence or absence of v-Src or EJ Ras as indicated. CAT activity is presented as a column graph and is given relative to that observed when the hybrid gene is cotransfected with the reporter. The degree of activation observed when the hybrid gene is cotransfected with the reporter plasmid into HepG2 cells is 15-fold that seen in L-M (TK⁻) cells. c, v-Src and EJ Ras enhance activation by a v-Jun:E2 hybrid in L-M (TK⁻) but not in HepG2 cells. An E2 reporter plasmid was transfected with v-Jun:E2 in the presence or absence of the v-src or *ras* genes as indicated. CAT activity relative to that measured when the hybrid gene was transfected with the reporter is given. The degree of activation observed when v-Jun:E2 is cotransfected with the reporter plasmid into HepG2 cells is 40-fold that seen in L-M (TK⁻) cells.

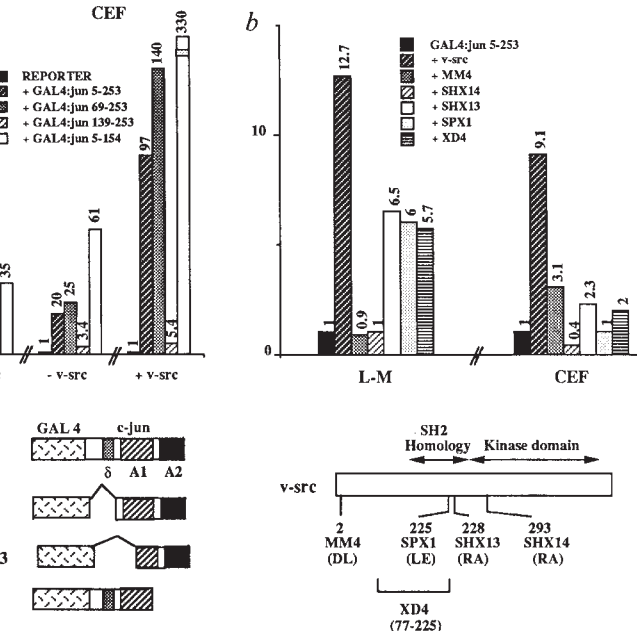
METHODS. L-M (TK⁻) and HepG2 cells were transfected similarly⁹. In b and c the medium was changed to DMEM supplemented with 0.5% fetal bovine serum 16–20 h after transfection. The stimulatory effects of v-Src and EJ Ras were greater when cells were propagated in growth medium supplemented with 0.5% rather than 10% serum following transfection (data not shown). The structure and origin of the reporter plasmids and the hybrid genes has been described⁹. v-Jun:E2 contains the DNA binding domain of the E2 protein from bovine papilloma virus in place of the Jun DNA binding domain. Five and one micrograms respectively of the GAL4 reporter and the GAL4:Jun hybrid plasmids and 2 and 3 μ g respectively of the E2 reporter and v-Jun:E2 hybrid plasmids were used for the transfections. pUCEJ6.6 (ref. 17) and



fpgv-1v-src (ref. 12) the plasmids containing respectively the EJ *ras* and v-src genes have been described and 2 μ g of these plasmids were used for transfection. CAT activity was measured as described⁹. The data presented in b and c is the average of two transfections each done in duplicate.

FIG. 2 a, The A1 domain of c-Jun is essential for enhancement of activation by v-src in L-M (TK⁻) cells and primary chicken embryo fibroblasts (CEF). Both cell types were transfected with a GAL4 reporter plasmid together with the indicated GAL4:Jun hybrids and either a vector lacking Src sequences (-Src) or one containing them (+v-Src). CAT activity is given relative to that observed when the reporter plasmid was transfected without any GAL4:Jun hybrid. The absolute CAT activity was different for the two cell types. CAT activity of the reporter plasmid transfected with v-Src was 3.2- and 1.8-fold higher in L-M (TK⁻) and CEF respectively than with reporter transfected alone. **b**, The capacity of v-Src mutants to enhance activation of GAL4:Jun 5-253 in L-M (TK⁻) and CEF cells. Both cell types were cotransfected with a GAL4 reporter plasmid and GAL4:Jun 5-253 and a vector lacking Src sequences, fpgv-1, v-Src, or the various src mutants indicated. CAT activity is relative to that observed when the GAL4 reporter and GAL4:Jun hybrid were transfected with fpgv-1. The bottom part of the figure shows the v-Src protein and the position of the changes in the mutant proteins¹². The amino acids deleted or inserted are indicated in parenthesis. MM4 has an insertion after the initiator methionine and is defective for myristylation (E. Liebl and G. S. Martin, personal communication). SHX14 has a linker inserted after amino acid 293 and lacks tyrosine kinase activity. SPX1 and SHX13 have insertions after amino acid 225 and 228 respectively. XD4 is deletion of amino acids 77-225. SHX13, SPX1 and XD4 all alter the SH2 domain in v-Src.

METHODS. CEF were cultured and transfected as described¹². One million primary CEF were plated on a 100-mm-diameter tissue-culture dish a day before transfection. Cells were collected 40 to 50 h after transfection. Five,



one and one micrograms respectively of the GAL4 reporter plasmid, GAL4:Jun hybrid and v-Src, or mutant v-Src plasmids were used for transfection. Two micrograms of the v-Src plasmids were transfected for L-M (TK⁻) cells and 16 to 20 h after transfection cells were transferred to medium containing 0.5% serum. The data presented represent the average of two transfections each done in duplicate. The mutant src genes used have all been described¹². GAL4:Jun 5-154 was constructed by deleting a *SacII* fragment from GAL4:Jun 5-253. The other GAL4:Jun hybrids have been described⁹.

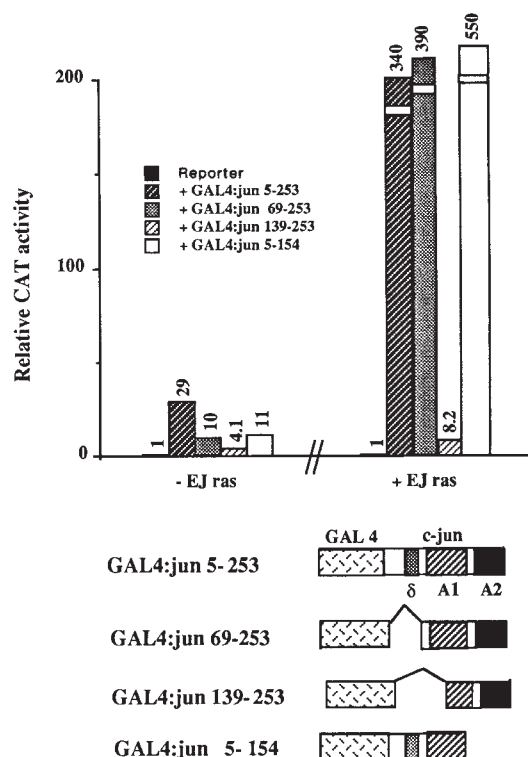
of EJ Ras also establishes that the Ras-induced increase in transcriptional activity of the GAL4:Jun hybrids is not merely the result of a general stimulation of transcription in these cells by the transfected *ras* gene. This stimulatory effect of Ras is particularly important because *c-jun* and an activated *c-Ha-ras* gene can cooperate in transforming primary rat embryo fibroblasts¹⁵; part of the cooperative effect might be the result of Ras increasing the transcriptional and transforming activity of Jun by decreasing the interaction between c-Jun and its inhibitor.

The results presented here demonstrate that the DNA binding and dimerization domains of c-Jun are not required for the stimulatory effect of Ras or Src and that an activation domain from c-Jun is probably sufficient for the effect. A model incorporating our results along with previous findings concerning the regulation of AP-1 activity is presented in Fig. 4. We propose that signal-transduction pathways which are triggered by v-Src or Ras expression lead to an increase in c-Jun transcriptional

activity. As the stimulatory effect of these oncogenes is not observed in a cell line lacking the c-Jun inhibitor we believe that this increase in c-Jun transcriptional activity results from disruption of the c-Jun:inhibitor interaction. This increased c-Jun transcriptional activity could in turn be amplified by an increase in c-Jun mRNA and protein levels, because the *c-jun* gene is positively autoregulated¹⁶. In this manner an increase

FIG. 3 The A1 domain of c-Jun is essential for enhancement of transactivation by EJ Ras in NIH3T3 cells. Five micrograms of the GAL4 reporter plasmid was cotransfected with 1 µg of the various GAL4:Jun hybrids and carrier DNA (-EJ Ras) or 1 µg of the EJ gene (+EJ Ras). CAT activity shown in each case is relative to that detected when the reporter plasmid was transfected alone. CAT activity of the reporter plasmid transfected with EJ Ras was 3.5-fold higher than that transfected alone. The inhibitory effect of the δ region is detected in chimaeric Jun molecules only when the heterologous DNA binding and dimerization domain is at the C terminal of the molecule, as in the JunE2 hybrids⁹ (Fig. 2c). In the GAL4:Jun hybrids used in this experiment the GAL4 DNA binding and dimerization domain is located at the amino terminus of the molecule and interferes with the inhibitory effect of the δ region (compare GAL4:Jun 69-253 to GAL4:Jun 5-253 or GAL4:Jun 5-154, in Fig. 5 (ref. 9), Fig. 2a and here). Changing the location of the GAL4 DNA binding and dimerization domain from the N terminus to the C terminus generates a chimaeric Jun:GAL4 molecule in which the inhibitory effect of the δ region is active (data not shown).

METHODS. Transfection of NIH3T3 cells was as described for L-M (TK⁻) cells. After transfection cells were grown in medium containing 0.5% serum. The data shown is the average of two transfections each done in duplicate.



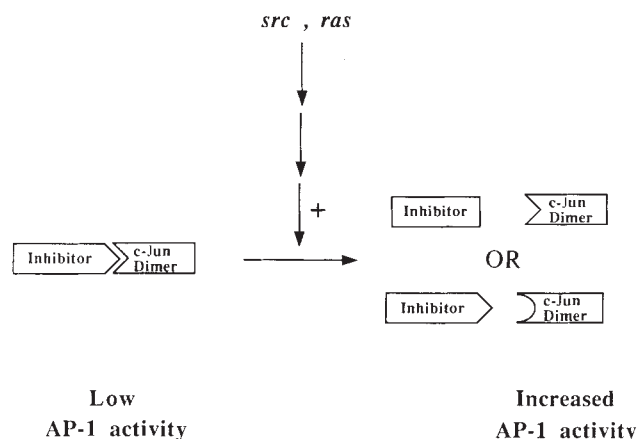


FIG. 4 A model for the increase in AP-1 activity induced by Src and Ras. The c-Jun protein is shown complexed with its inhibitor which reduces the transcriptional activity of c-Jun. The Src and Ras proteins are shown initiating a cascade of events that results in disruption of the c-Jun:inhibitor complex and an increase in c-Jun transcriptional activity. Two possibilities are considered. In one, the signal transduction pathway leads to a modification of the inhibitor and makes it incompetent to interact with c-Jun. Alternatively, the pathway could lead to modification of c-Jun allowing it to escape regulation by the inhibitor. A single pathway is shown to mediate the effect of Src and Ras for simplicity. The observation that the SH2 domain in Src may be involved in its action suggests that part of the effect of Src is through the GAP/p62 pathway and through Ras^{13,14}. It is possible that many different pathways are operative for regulating AP-1 activity, some even including other components of AP-1, for example c-Fos.

in c-Jun transcriptional activity, even in the absence of changes in c-Fos, could induce AP-1 activity in response to extracellular stimuli. The present findings therefore emphasize the importance of the c-Jun inhibitor in mediating changes in c-Jun transcriptional activity in response to signalling events in the cell. In conjunction with the previous findings about the differential interaction of this inhibitor with ν - and c-Jun⁹, our results suggest that this inhibitor may represent a vital regulator of c-Jun activity. A corollary of this conclusion is that signal-transduction pathways starting from Src and Ras could enhance c-Jun transcriptional activity by modifying the inhibitor and preventing it from interacting with c-Jun, rather than acting on Jun directly. Thus these findings identify another level at which c-Jun activity is controlled and bring us one step closer to understanding the regulation of c-Jun and AP-1 activity in response to extracellular stimuli. It remains to be determined whether growth factors and other agents that regulate AP-1 activity also operate through the c-Jun inhibitor and the δ -A1 domain of c-Jun. □

Structure of the detoxification catalyst mercuric ion reductase from *Bacillus* sp. strain RC607

N. Schiering, W. Kabsch, M. J. Moore*, M. D. Distefano*, C. T. Walsh* & E. F. Pai†

Max-Planck-Institut für Medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, D-6900 Heidelberg, Germany

* Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, 240 Longwood Avenue, Boston, Massachusetts 02115, USA

SEVERAL hundred million tons of toxic mercurials are dispersed in the biosphere¹. Microbes can detoxify organo-mercurials and mercury salts through sequential action of two enzymes, organomercury lyase² and mercuric ion reductase (MerA)^{3–5}. The latter, a homodimer with homology to the FAD-dependent disulphide oxidoreductases⁶, catalyses the reaction $\text{NADPH} + \text{Hg(II)} \rightarrow \text{NADP}^+ + \text{H}^+ + \text{Hg(0)}$, one of the very rare enzymic reactions with metal substrates. Human glutathione reductase^{7,8} serves as a reference molecule for FAD-dependent disulphide reductases and between its primary structure⁹ and that of MerA from *Tn501* (*Pseudomonas*), *Tn21* (*Shigella*), p1258 (*Staphylococcus*) and *Bacillus*, 25–30% of the residues have been conserved^{10,11}. All MerAs have a C-terminal extension about 15 residues long but have very varied N termini. Although the enzyme from *Streptomyces lividans* has no addition, from *Pseudomonas aeruginosa* *Tn501* and *Bacillus* sp. strain RC607 it has one and two copies respectively of a domain of 80–85 residues, highly homologous to MerP, the periplasmic component of proteins encoded by the *mer* operon¹¹. These domains can be proteolytically cleaved off without changing the catalytic efficiency³. We report here the crystal structure of MerA from the Gram-positive bacterium *Bacillus* sp. strain RC607. Analysis of its complexes with nicotinamide dinucleotide substrates and the inhibitor Cd(II) reveals how limited structural changes enable an enzyme to accept as substrate what used to be a dangerous inhibitor. Knowledge of the mode of mercury ligation is a prerequisite for understanding this unique detoxification mechanism.

The three-dimensional structure of MerA from *Bacillus* sp. strain RC607 can be clearly divided into three parts: the two N-terminal regions (residues 1–166), the core (167–616), the equivalent of the glutathione reductase structure, and the C-terminal extension (residues 617–631). This partition is also reflected in the electron-density map (Table 1). The two regions at the N terminus cannot be modelled owing to the considerable mobility of this part of the structure. Although the electron density corresponding to the C-terminal tail does allow tracing of the chain it is less well defined than that for the core of the molecule. Most of this structure, which is common to MerA and glutathione reductase, adopts a very similar fold in both proteins (Fig. 1). There are, however, some local differences. They affect residues 229–240, which connect the two long α -helical protrusions of the FAD domain. The first of these helices is shorter than its counterpart in glutathione reductase, missing the terminal 3_{10} -helical turn. The connecting loop is located closer to the interface region; this results in a shift of helix α_9 by about 7 Å. Residues 237–241 form an antiparallel β -pleated sheet with their symmetrically equivalent partners of the second subunit. Whereas in glutathione reductase the corresponding region is the most mobile of the whole protein chain⁷, the electron density indicates a rather rigid structure in MerA. Another small but interesting difference between MerA and glutathione reductase is the insertion of Lys 263 in the middle of helix α_3 . This results in the widening of one turn into a π

Received 28 February; accepted 21 May 1991.

- Curran, T. & Franza Jr, B. R. *Cell* **55**, 395–397 (1988).
- Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741–752 (1987).
- Angel, P. et al. *Cell* **49**, 729–739 (1987).
- Ryder, K., Lanahan, A., Perez-Albuern, E. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1500–1503 (1989).
- Lamph, W. W., Wamsley, P., Sassone-Corsi, P. & Verma, I. M. *Nature* **334**, 629–631 (1988).
- Ryseck, R.-P., Hirai, S. I., Yaniv, M. & Bravo, R. *Nature* **334**, 535–537 (1988).
- Wasylyk, C., Imier, J. L. & Wasylyk, B. *EMBO J.* **8**, 2475–2483 (1988).
- Quartin, B. & Breathnach, R. *Nature* **334**, 538–539 (1988).
- Baichwal, V. R. & Tjian, R. *Cell* **63**, 815–825 (1990).
- Bohmann, D. & Tjian, R. *Cell* **59**, 709–717 (1989).
- Bos, T. J. et al. *Genes Dev.* **4**, 1677–1687 (1990).
- DeClue, J. D. & Martin, G. S. *J. Virol.* **63**, 542–554 (1989).
- Cantley, L. C. et al. *Cell* **64**, 281–302 (1991).
- Moran, M. F. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8622–8626 (1990).
- Schutte, J., Minna, J. D. & Birrer, M. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2257–2261 (1989).
- Angel, P., Hattori, K., Smeal, T. & Karin, M. *Cell* **55**, 875–885 (1988).
- Shih, C. & Weinberg, R. A. *Cell* **29**, 161–169 (1982).

ACKNOWLEDGEMENTS. We thank the Martin laboratory for providing primary chicken embryo fibroblasts and the v-src plasmids, S. Martin, E. Liebl and L. England for helpful discussions, members of our laboratory for reading the manuscript and W. Sugden for his comments. V.R.B. was partly supported by a junior postdoctoral fellowship of the American Cancer Society. This work was funded by the NIH.

† To whom correspondence should be addressed.

TABLE 1 Statistics of data collection and refinement

DATA COLLECTION		Soak conditions					
	Native	0.5 mM EMP, 35 h	Saturated Pten ₂ Cl ₂ +1 mM UO ₂ Ac ₂ , 10 days	1) 5 mM NADH, 30 min 2) 1 mM HgCl ₂ , 16 h	10 mM NADPH	2.5 mM NADP (cocrySTALLIZED)	5 mM NADPH + 1.5 mM Cd acetate 4.5 h
Resolution (Å)	3.0	3.4	3.5	4.0	3.0	3.0	3.3
Observed reflections (no.)	111,480	123,132	145,561	94,110	92,966	87,318	42,889
Unique reflections (no.)	25,085	19,546	21,029	16,331	18,818	18,542	13,257
Completeness (%)	98	99	100	99	87	87	82
R _{sym} (%)*	10.1	9.5	11.6	14.6	11.6	10.5	19.6
R _i (%)†		16.9	21.1	22.6			
REFINEMENT							
	Resolution (Å)	100–6.0	6.0–4.5	4.5–3.5	Total		
	⟨m⟩‡	0.59	0.56	0.43	0.50		
EMP	R _C §	60.7	67.1	76.0	66.9		
	F _H /E¶	1.6	1.5	1.1	1.31		
Pten ₂ Cl ₂ + UO ₂ Ac ₂	R _C	65.0	74.5	84.4	73.4		
	F _H /E	1.5	1.2	0.8	1.09		
HgCl ₂	R _C	64.5	73.4	82.1	71.6		
	F _H /E	1.4	1.4	0.9	1.14		

Tetragonal crystals (I422, $a=b=180.5$ Å, $c=127.5$ Å) of MerA from *Bacillus* sp. strain RC607 were obtained as described²³. For data collection and soaking experiments involving heavy atoms the crystals were transferred into 1.5 M LiSO₄, 50 mM KP_i, pH 7.0. Measurements of NADPH complexes were done with the crystals mounted in buffer-filled capillaries to avoid reoxidation. X-rays (CuK_α) were generated by a GX-18 rotating anode (Elliot/Enraf-Nonius, Delft) operated at 35 kV and 50 mA and focused by Francks double-mirror optics. Diffraction data were recorded at about 4 °C by an electronic area detector (Siemens/Nicolet, Madison, Wisconsin) and evaluated by the program XDS^{24,25}. Heavy atom positions in the derivatives were found by deconvolution of the difference Pattersons; minor sites were obtained from difference Fourier syntheses and refined subsequently²⁶. An initial MIR electron-density map was calculated and slightly improved by solvent flattening. There is one monomer in the asymmetric unit (67 % solvent), correcting earlier assumptions²³. With a model of glutathione reductase⁷ as a guide, residues 167–615 were fitted to the electron density using the interactive graphics program FRODO²⁷ adapted to an IRIS 4GT (Silicon Graphics, Mountain View, California) by C. M. Cambillau. Electron density representing the two N-terminal domains is not sufficiently well defined to allow chain tracing. The resulting model was refined with the program XPLOR²⁸. We adopted the standard protocol for heating, cooling, minimization and calculation of overall temperature factors. A new map was calculated combining information from the model with the derivative information. Model bias was reduced by incorporating the method developed by Read²⁹. After another round of rebuilding and refinement, the *R* factor dropped to 23.9 % including all data from 10 to 3 Å resolution and using an overall temperature factor of 18 Å². The r.m.s. deviations of the model from ideal bond length, bond angles and planarity are 0.025 Å, 4.1° and 1.7°, respectively. There are <10 non-glycine residues in the 'disallowed' regions of a Ramachandran plot. The map for the NADPH complex was calculated using the native model combined with the derivative information for phase calculation. The NADP⁺ and Cd data sets were processed accordingly. Observing good geometry (NADPH: 0.025 Å (r.m.s. deviations for bond lengths), 4.2° (angles), 1.9° (planarity); NADP⁺: 0.026 Å, 4.3°, 1.9°; Cd: 0.026 Å, 4.6°, 2.3°), refinement converged to *R* factors of 22.3 %, 22.4 % and 21.0 % (corresponding constant *B* factors: 15 Å², 16 Å², 14 Å²) for the NADPH, NADP⁺ and Cd complexes, respectively. All data from 10 Å to 3.0 Å resolution (NADPH, NADP⁺) and 10 Å to 3.3 Å (Cd) were included. In addition, the binding site of the Cd ion was confirmed by difference Fourier techniques. Abbreviations, EMP, ethylmercury phosphate; Pten₂Cl₂, platinum diethylenediamine dichloride; UO₂Ac₂, uranyl acetate.

* $R_{\text{sym}} = \sum_i \sum_j |I_{h,j} - I_{h,i}| / \sum_i \sum_j I_{h,j}$, where $I_{h,j}$ are symmetry related intensity observations and $I_{h,i}$ is the mean intensity of reflections with unique indices h .

† $R_i = 2\sum_h |F_{\text{PH}} - F_P| / \sum_h |F_{\text{PH}} + F_P|$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively.

‡ $\langle m \rangle$ = mean figure of merit.

§ $R_C = \sum |F_{\text{PH,obs}} - F_{\text{PH,calc}}| / \sum |F_{\text{PH,obs}} + F_P|$ for centric reflections.

¶ $F_H/E = \sqrt{\sum f_H^2 / \sum (F_{\text{PH,obs}} - F_{\text{PH,calc}})^2}$.

helix and a slight bend in the helix axis; also affected by this insertion is Tyr 264.

FAD binds to MerA in the same extended conformation as in glutathione reductase. Most interactions between protein and prosthetic group have been conserved. When the product NADP⁺ is bound to the enzyme, only the 2'-P-5'-AMP part of the dinucleotide is represented by electron density, whereas the NMP half is flexible. The main chain of residues 340–344 moves up to about 2 Å, mainly to make room for the phosphate groups. The guanidinium group of Arg 365 is roughly parallel to the adenine base and, together with Ser 366, binds tightly to the 2'-phosphate, which replaces a phosphate or sulphate ion bound there in the holoenzyme.

The substrate NADPH lies in a narrow pocket on the surface of MerA, adopting an extended conformation as in glutathione reductase¹². On binding of NADPH we observe movements slightly larger than those described for NADP⁺ (Fig. 2). In addition, the side chain of Tyr 344, which in the substrate-free enzyme points perpendicularly onto the flavin ring, has swung away, opening a crevice and allowing the nicotinamide to stack on the *re*-face of the isoalloxazine ring. The hydroxyl of the tyrosine has moved by about 7 Å from its original position. Its aromatic ring provides a backstop on the *re*-side of the

nicotinamide. The distance between the ring systems of the substrate and the flavin is suitable for the transfer of hydride equivalents. Next to C4 of the nicotinamide, there is a salt bridge between Lys 216 and Glu 348, perfectly placed to assist in catalysis (Fig. 3).

Unlike the other disulphide oxidoreductases, which cycle between their fully oxidized E_{ox} and two-electron reduced EH₂ forms during catalysis, MerA is thought to use its EH₂ and four-electron reduced EH₄ forms¹³. Although the binding mode and site of NADPH in MerA are very similar to those in glutathione reductase, the area around the *si*-side of the flavin ring has been modified. There is both a reducible active-site disulphide as in the homologous enzymes and, uniquely, a geminal cysteine pair close to the C terminus, required for catalysis in a proposed intersubunit active-site arrangement¹⁴. In the oxidized structure, modelling the cysteine residues as disulphides agrees well with the observed electron density. The redoxactive disulphide Cys 207 to Cys 212 is at the beginning of the long helix α 2 of the FAD domain next to atom C4a of the flavin.

On interaction with NADPH both disulphides are reduced^{15,16}. The proximal Cys 212 forms a charge transfer complex with flavin and the sulphur of the distal Cys 207 moves

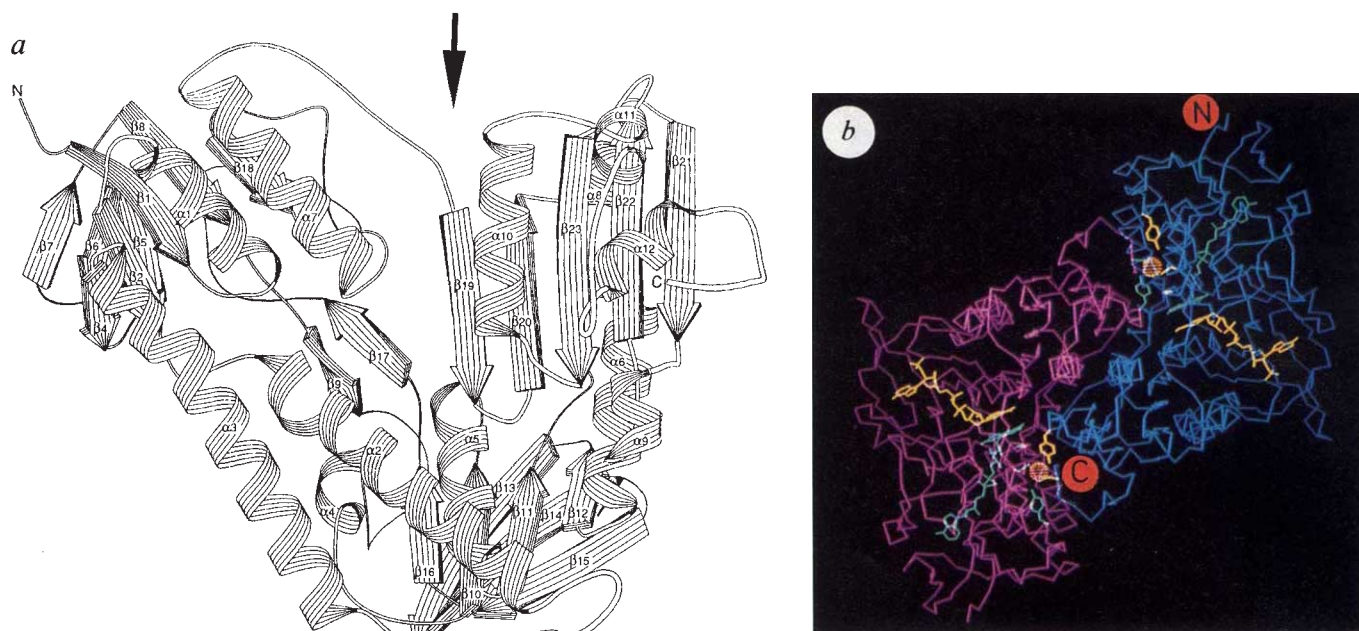


FIG. 1 a, Ribbon representation³⁰ of residues 167–630 of the MerA monomer (α = α helix; β = β sheet). Secondary structure has been assigned using the program DSSP³¹. The topology of the core (residues 167–616) is the same as that in human glutathione reductase⁸. Residues 167–305 represent the FAD domain, residues 306–434 the NADP, residues 435–504 the central and residues 505–630 the interface domain, which in MerA includes the short C-terminal tail. Of the α positions, 340 can be superimposed³² for MerA and glutathione reductase with an r.m.s. deviation of 0.97 Å, the closest fits are in the interior parts of the domains. The relative positions of the domains are not changed significantly. This is different from what has been found for lipoamide dehydrogenase³³, where the domains move

and quite contrary to the results obtained with thioredoxin reductase, where entire domains are missing and others are rotated by more than 60° (ref. 34). The arrow on top indicates the direction of view of b. b, Alpha-carbon trace of the MerA dimer. View perpendicular to the one shown in a, with the molecule rotated by 90° towards the viewer. For both subunits bound FAD (green) and NADPH molecules (yellow) and the Cd ions (dot surfaces) with their ligands are indicated, as are the N and C termini of the blue subunit. Each of the two symmetry-related metal binding sites is built up from side chains of both subunits (Cys 207, Tyr 264, Tyr 605', Cys 628' and for the second site Cys 207', Tyr 264', Cys 628, Tyr 605).

towards the side chain of Cys 628'. Cys-629' adopts a conformation in which its side chain points further away from the active site. The distance between the two sulphurs of Cys 207 and Cys 628' is 5 Å, making them good candidates for metal-ion ligation.

Besides these two residues we have identified two further candidates, Tyr 264 and Tyr 605', for binding of the metal substrate. Altering the straight geometry of helix α 3 by inserting a lysine next to Tyr 264, the hydroxyl of this amino acid is shifted 2 Å closer to the sulphur atom of Cys 207 than is found for the

corresponding residues in glutathione reductase, Tyr 114 and Cys 58 (ref. 7). At a distance of 5.8 Å from this oxygen we find another hydroxyl group (Fig. 3). It belongs to Tyr 605', which in MerA has replaced the catalytic histidine (His 467' in glutathione reductase), which in the disulphide reductases has been assigned an essential mechanistic role⁴. The lines connecting the hydroxyls of Tyr 264 and Tyr 605' and the sulphurs of Cys 207 and Cys 628', respectively, are close to perpendicular but offset by 1.3 Å.

Steady-state studies indicate a finely balanced role of

FIG. 2 Conformational changes caused by the binding of NADPH (colour-coded) and involving residues 340 to 344. The movement of this section from its old (orange) to its new position (green) can be described by a 48° rotation around an axis connecting atoms C340 and C α 344. This leads to a maximum displacement of about 3 Å for the C α atom of Ser 342, more pronounced than the corresponding change described in glutathione reductase³⁵. For the first time in the glutathione reductase-related enzyme family, kinetic cooperativity was observed with the MerA from Tn501 (ref. 19). The authors of ref. 19 proposed an alternating sites model, where binding at one active site is coupled to catalytic conversion at the other and the active sites reverse roles after completion of a catalytic cycle. Unfortunately, the only crystal form we have obtained so far has a monomer in the asymmetric unit; this is also true for crystals grown with NADP⁺ in the solution. We cannot, therefore, comment on the conclusions of ref. 19 as crystal symmetry will superimpose eventual differences between subunits.

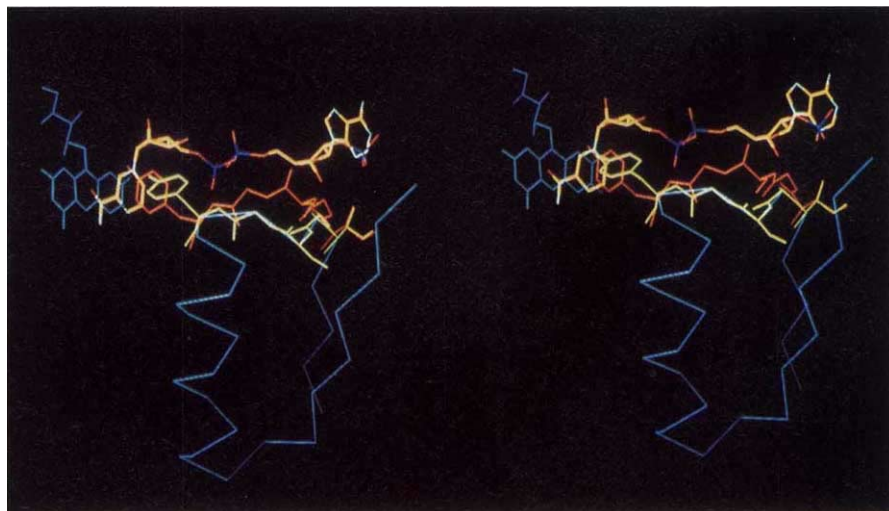
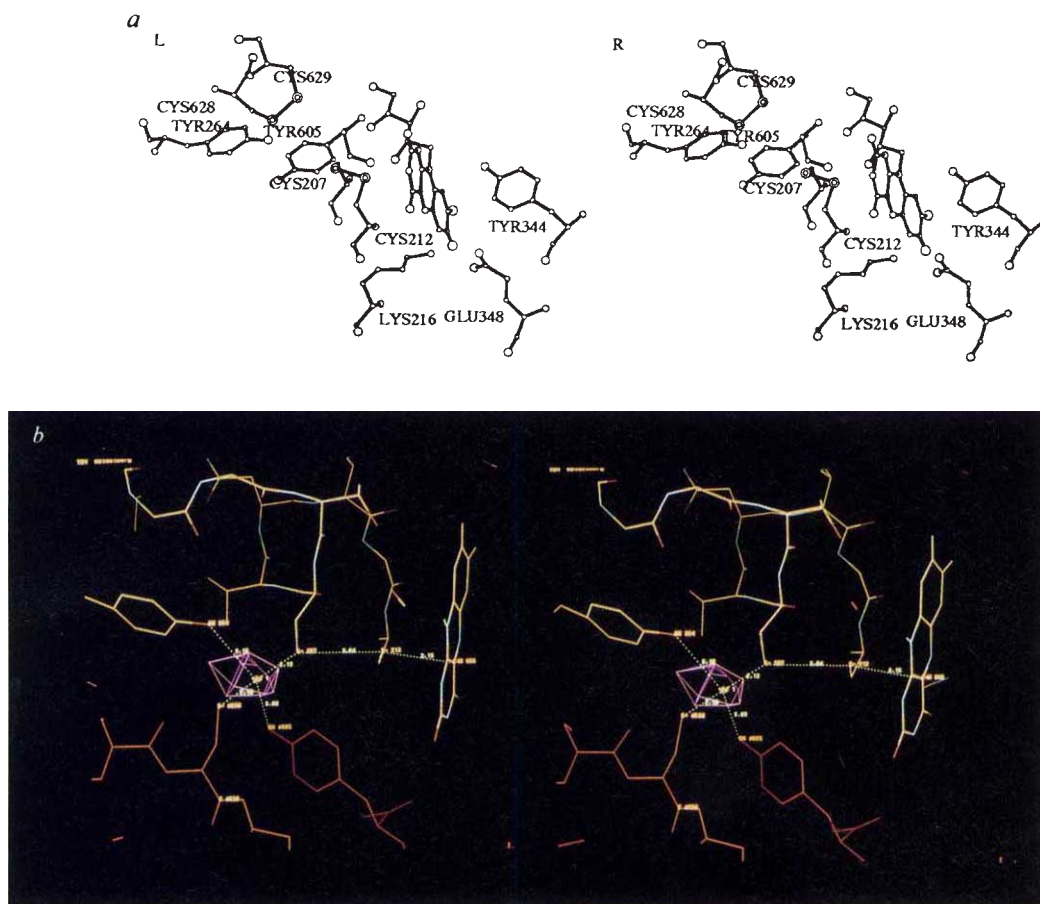


FIG. 3 *a*, Stereo view of the active site arrangement in the oxidized, substrate-free enzyme. Tyr 605, Cys 628 and Cys 629 belong to the second subunit. *b*, Stereo representation of the peak obtained by difference Fourier analysis of crystals soaked in 5 mM NADPH and 1.5 mM cadmium acetate. The contours (pink) are drawn at 60 % of the maximum difference electron density. Possible ligands are the thiolate of Cys 207 and the hydroxyl of Tyr 264 (both colour-coded) as well as Tyr 605' and Cys 628' (both red) from the other subunit. To avoid or minimize overlap, the view is rotated by approximately 90° towards the viewer from the orientation of *a*.



exogenous thiols in the reaction catalysed by MerA. The thiols must be present in sufficient quantity to prevent formation of an inhibitory complex, but they inhibit the reaction when in excess¹⁷. We have not yet identified conditions that would allow binding of Hg(II) to the active site and still keep the crystals structurally intact, which is possibly due to the proposed co-operative behaviour¹⁸, which in our crystal form would lead to the breaking of crystal symmetry. Catalytically incompetent mutants of the *Bacillus* enzyme are being constructed, which should allow cocrystallization with Hg(II) and reduced pyridine nucleotides.

To assess, nevertheless, whether this locale was the binding site for Hg(II), we turned to soaking with Cd(II), a non-reducible analogue of Hg(II), known to inhibit Hg(II) reduction¹⁹ and to coordinate to the active site²⁰. Soaking crystals in a solution containing NADPH and cadmium acetate resulted in clean maps with additional electron density surrounded by the distorted tetrahedron of the hydroxyls of Tyr 264 and Tyr 605' and the sulphurs of Cys 207 and Cys 628' (Fig. 4). This peak was only observed when pyridine nucleotide reduced enzyme was exposed to Cd(II). If a cadmium ion is placed in this density the liganding cysteine sulphurs are 2.2 Å away and the oxygens of the tyrosines 3.1 Å. Although this is not an active complex, it is suggestive of at least one mode of Hg(II) ligation in the active site. Our results are most consistent with a mechanism involving a C4a adduct⁵, the facile formation of which has been described^{15,21}. The orange colour of the crystals of the NADPH-Cd(II) complex is an indication that the charge-transfer complex is still intact. We are, therefore, dealing with a reaction state different from the one previously analysed by EXAFS (ref. 22).

These results give the first picture of structural organization of this unique Hg(II) reduction and detoxification enzyme. They reveal the placement of the four thiols of consequence in catalysis. The Cd(II) complex results suggest that Hg(II), in addition to its location between the oxygens of Tyr 264 and Tyr 605', may

be in an intersubunit bidentate sulphur ligation between Cys 628' and Cys 207. Both cysteines are liberated on reduction of the enzyme by NADPH. It is not obvious how a tri- or tetradentate sulphur complex to Hg(II) or Cd(II) could form without significant structural reorganization. The Cd(II) binding results suggest specific mutations to assess some of these issues. □

Received 25 March; accepted 20 May 1991.

- Summers, A. O. & Silver, S. A. *Rev. Microbiol.* **32**, 637-672 (1978).
- Begley, T. P., Walts, A. E. & Walsh, C. T. *Biochemistry* **25**, 7186-7192 (1986).
- Fox, B. & Walsh, C. T. *J. biol. Chem.* **257**, 2498-2503 (1982).
- Misra, T. K. *et al. Gene* **34**, 253-262 (1985).
- Moore, M. J., Distefano, M. D., Zydowsky, L. D., Cummings, R. T. & Walsh, C. T. *Acc. chem. Res.* **23**, 301-308 (1990).
- Williams, C. H. Jr *The Enzymes* **13**, 90-165 (1976).
- Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **195**, 701-729 (1987).
- Pai, E. F. & Schulz, G. E. *J. biol. Chem.* **258**, 1752-1757 (1983).
- Krauth-Siegel, R. L. *et al. Eur. J. Biochem.* **121**, 259-267 (1982).
- Laddaga, R. A., Chu, L., Misra, T. K. & Silver, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5106-5110 (1987).
- Wang, Y. *et al. J. Bacteriol.* **171**, 83-92 (1989).
- Pai, E. F., Karplus, P. A. & Schulz, G. E. *Biochemistry* **27**, 4465-4474 (1988).
- Distefano, M. D., Au, K. G. & Walsh, C. T. *Biochemistry* **28**, 1168-1183 (1989).
- Distefano, M. D., Moore, M. J. & Walsh, C. T. *Biochemistry* **29**, 2703-2713 (1990).
- Miller, S. M. *et al. Biochemistry* **29**, 2831-2841 (1990).
- Miller, S. M., Moore, M. J., Massey, V., Williams, C. H. Jr & Walsh, C. T. *Biochemistry* **28**, 1194-1205 (1989).
- Miller, S. M., Ballou, D. P., Massey, V., Williams, C. H. Jr & Walsh, C. T. *J. biol. Chem.* **261**, 8081-8084 (1986).
- Miller, S. M., Massey, V., Williams, C. H. Jr, Ballou, D. P. & Walsh, C. T. *Biochemistry* **30**, 2600-2612 (1991).
- Rinderle, S. J., Booth, J. E. & Williams, J. W. *Biochemistry* **22**, 869-876 (1983).
- Distefano, M. D. thesis, Massachusetts Institute of Technology (1989).
- Sahlmann, L., Lambeir, A.-M. & Lindsag, S. *Eur. J. Biochem.* **156**, 479-488 (1986).
- Raybuck, S. A., Distefano, M. D., Teo, B.-T., Orme-Johnson, W. & Walsh, C. T. *J. Am. chem. Soc.* **112**, 1983-1989 (1990).
- Moore, M. J., Distefano, M. D., Walsh, C. T., Schiering, N. & Pai, E. F. *J. biol. Chem.* **264**, 14386-14388 (1989).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 916-924 (1988).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 67-71 (1988).
- Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. *Acta crystallogr.* **B24**, 997-1003 (1968).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Read, R. J. *Acta crystallogr.* **A42**, 140-149 (1986).
- Priestle, J. P. *J. appl. Crystallogr.* **21**, 572-576 (1988).

31. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
 32. Kabsch, W. *Acta crystallogr.* **A34**, 827–828 (1978).
 33. Mattevi, A., Schierbeek, A. J. & Hol, W. G. J. *J. molec. Biol.* (in the press).
 34. Kuriyan, J. *et al.* *Nature* **352**, 172–174 (1991).
 35. Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **210**, 163–180 (1989)

ACKNOWLEDGEMENTS. We thank K. Fritz-Wolf for help in data collection, G. A. Petsko for discussions and K. C. Holmes for support. This work was supported by the Deutsche Forschungsgemeinschaft and the National Institutes of Health.

Convergent evolution of similar function in two structurally divergent enzymes

John Kuriyan^{*†}, T. S. R. Krishna^{*†}, Lim Wong^{*},
 Brian Guenther^{*}, Arno Pahler^{‡§},
 Charles H. Williams Jr^{||} & Peter Model^{*}

^{*} The Rockefeller University, 1230 York Avenue, New York 10021, USA

[†] Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York 10021, USA

[‡] Department of Molecular Biophysics and Biochemistry, Columbia University, 630 W 168th Street, New York 10032, USA

^{||} Department of Veterans Affairs Medical Center and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48105, USA

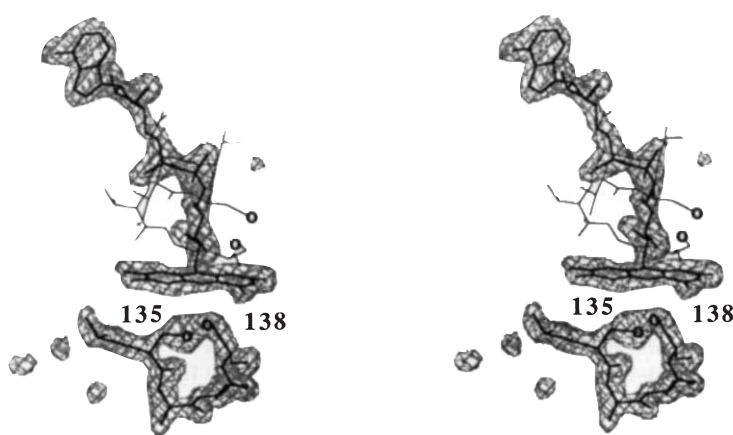
[§] Present address: First Department, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

AN example of two related enzymes that catalyse similar reactions but possess different active sites is provided by comparing the structure of *Escherichia coli* thioredoxin reductase with glutathione reductase¹. Both are dimeric enzymes that catalyse the reduction of disulphides by pyridine nucleotides through an enzyme disulphide and a flavin². Human glutathione reductase contains four structural domains within each molecule: the flavin-adenine dinucleotide (FAD)- and nicotinamide-adenine dinucleotide phosphate (NADPH)-binding domains, the 'central' domain and the C-terminal domain that provides the dimer interface and part of the active site^{3,4}. Although both enzymes share the same catalytic

mechanism and similar tertiary structures, their active sites do not resemble each other^{5,6}. We have determined the crystal structure of *E. coli* thioredoxin reductase at 2 Å resolution, and show that thioredoxin reductase lacks the domain that provides the dimer interface in glutathione reductase, and forms a completely different dimeric structure. The catalytically active disulphides are located in different domains on opposite sides of the flavin ring system. This suggests that these enzymes diverged from an ancestral nucleotide-binding protein and acquired their disulphide reductase activities independently.

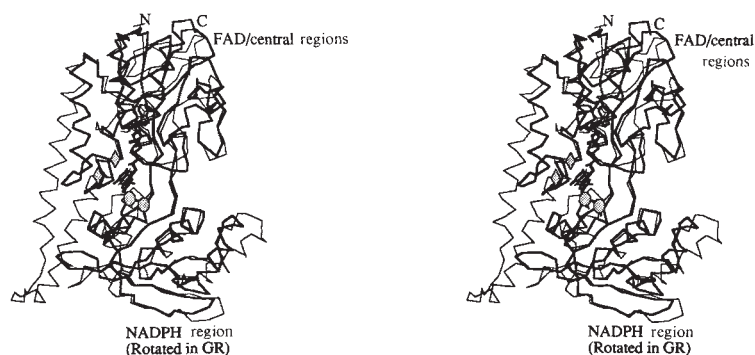
Thioredoxins are small redox-active proteins that have diverse functions. In *E. coli*, thioredoxin is reduced by NADPH-dependent thioredoxin reductase (TR; ref. 7). The three-dimensional structure of TR was determined by multiple isomorphous replacement (Fig. 1) and consists of two molecules which interact closely to form a symmetric dimer. Each molecule has three clearly delineated domains, which correspond to the FAD, NADPH and central domains of glutathione reductase (GR) (Fig. 2). The tertiary structure of TR is very similar to that of the first three domains of GR, the chief differences being a change in the orientation of the NADPH domain and the deletion of a helix (Fig. 2). Nineteen of the 21 β strands in TR correspond to elements in GR; the other two are located at the N terminus and near the redox-active disulphide that is a distinct feature of the TR sequence (Fig. 3). Likewise, five of the seven α helices and three of the four 3_{10} helices in TR align with corresponding elements in GR. A notable feature of the structural alignment is the relative displacement of the NADPH domains in the two enzymes. This corresponds to a rotation of one of these domains by 66° about the two β strands that connect it to the FAD and central domains, leaving the backbone structure relatively unperturbed (Fig. 2). On aligning the FAD/central domains and the NADPH domain separately, the root-mean-square (r.m.s.) deviation in C_{α} positions is 1.8 Å for the FAD/central domains and 2.0 Å for the NADPH domain for core residues comprising about 70% of the TR sequence. Including in the alignment only those C_{α} atoms that deviate by less than 2.5 Å results in r.m.s. deviations of 1.1 Å both in the FAD/central domains and in the NADPH domain, with 54% and 65% of the TR residues satisfying this criterion in the two parts of the structure, respectively. These results are similar to

FIG. 1 Stereo diagram of electron density at 2.0 Å resolution for the FAD group and redox-active disulphide of mutant TR (Cys 138 to Ser). The model for the FAD group and residues 134–138 in TR are shown in thick lines, with the sulphur and oxygen side-chain atoms of residues 135 and 138 indicated with open circles. The N^{10} atom of the flavin ring system is at the bottom. The model for residues 57–63 in GR¹ is also shown in thin lines. The C_{α} atoms of the GR FAD/central domains were optimally superimposed on TR (Figs 2 and 3). This transformation results in the FAD groups of the two proteins being closely aligned (r.m.s. deviation, 0.8 Å). The resulting orientation of the active-site residues 57–63 in GR are shown in thin lines, with the redox-active sulphurs of residues 58 and 63 indicated by open circles. The electron density was generated using coefficients $(|F_o| - |F_c|) \exp(i\alpha_c)$, where F_o and F_c are the observed and calculated structure factors and α_c is the calculated phase, with the FAD group and residues 134–138 omitted from the phase calculation. The final refined structure (Fig. 2) was used for the calculation. Contour levels corresponding to 3.5 s.d. above the mean are shown. Crystals of the (Cys 135 to Ser) mutant protein¹⁶, space group $P6_322$, with $a=b=123.7$ Å and $c=81.6$ Å (ref. 17), were soaked for 16–20 h in 40% polyethylene glycol ($M_r=3,500$) with 0.2 M lithium sulphate and either 2 mM ethyl mercury phosphate (EMP) or 0.2 mM 1,4-diacetoxy mercury-2,3-dimethoxybutane (Baker's dimercurial). An Enraf-Nonius FAST area detector system mounted on an Elliot GX-21 rotating anode X-ray generator was used to collect data sets to 3 Å resolution, using only one each of the native and derivative crystals. The resulting native data set is 95% complete to 3 Å resolution, and has an overall $R_{sym}(I)$ of 5%. Five mercury sites for the EMP derivative and two sets of paired sites for the Baker's dimercurial were



located by Patterson map and difference Fourier analysis. Refinement of heavy atom parameters¹⁸ resulted in phasing powers, f_i/E , of 1.37 and 1.27 for EMP and Baker's dimercurial, respectively, including anomalous scattering. The phases were further improved by solvent flattening¹⁹, increasing the mean figure of merit from 0.59 to 0.85. These phases were used to calculate a map at 3 Å resolution, from which the main chain was readily traced. Refinement at 3 Å resolution was carried out by restrained molecular dynamics using the programs X-PLOR^{20,21} and FRODO²². The resulting structure was used directly to initiate refinement at 2 Å resolution, using data for the (Cys 138 to Ser) protein (Fig. 2).

Fig. 2 Stereo diagram of the C α atoms of TR and the FAD/central domains and the rotated NADPH domain of GR. Thick lines, TR C α atoms (residues 1–316) and all nonhydrogen atoms of the FAD group. The last four residues are not visible in the electron density map. The active-site cysteines are indicated by shaded circles, below the flavin ring system. Thin lines, GR FAD (residues 18–158), central (291–356) and NADPH (159–290) domains, and the FAD group¹. The active-site cysteines are indicated by shaded diamonds, above the flavin. The FAD/central (taken together) and the NADPH domains of GR were superimposed on the corresponding domains of TR by rigid-body rotations and translations (Fig. 3). Once the FAD/central domains of the two proteins were brought into alignment, the NADPH domain of GR was rotated by 66° as a rigid body in order to align it with the NADPH domain of TR. This motion can be described as a rotation about the two-stranded β -sheet separating the regions, with atoms in the hinge region (residues 160 and 289) shifting only by ~1.5 Å. The helix formed by residues 63–80 in GR is missing in TR. The structural model of TR is for a (Cys 138 to Ser) mutant. Data to 2 Å resolution were measured at Beamline X12-C, Brookhaven National Laboratory (in collaboration with R. M. Sweet). The current model for this mutant includes 2,447 non-hydrogen protein and FAD atoms and 217 water molecules, with individual temperature factors refined for all atoms. The *R*-factor is 17.7% for data between 6 Å and 2.0 Å, including 21,106 observed structure factors with amplitudes greater than 2 σ , which corresponds to 93% of the unique data. The deviations from ideal bond lengths and angles are 0.0019 Å and 3.2°, respectively, and



only three non-glycine residues have backbone dihedral angle values in disallowed regions of the Ramachandran diagram²³. The structures of the two TR mutants are very similar, with r.m.s. deviations of backbone atoms of 0.1 Å and with the terminal sulphur and oxygen atoms of residues 135 and 138 in positions that are indistinguishable at this resolution. This suggests that these structures provide good models for that of the wild-type enzyme. Details of the structure and refinement at 2 Å resolution will be published separately (T.S.R.K., R. M. Sweet, C.H.W. and J.K., manuscript in preparation). The structural model for GR is that of Karplus and Schulz, refined at 1.54 Å resolution (Protein Databank Entry 3GRS)¹.

the deviations obtained on superimposing globin structures with comparable sequence identity⁸.

The redox-active cysteines are in different locations in the two enzymes (Figs 1 and 2). But, both enzymes have sequence or structural similarities in the regions corresponding to the disulphide of the other enzyme. The sequence TSDGFF (single-letter amino-acid notation) in GR (residues 176–181) aligns with

the active-site sequence TCDGFF in TR, and this hexapeptide is the one with greatest sequence similarity in the entire alignment (Fig. 3). At the site of the GR disulphide (residues 58–63), the first turn of the α helix is in structural alignment with a 3_{10} helix in TR (Figs 2 and 3). Together with the homology in secondary structural elements and the presence of several other clustered regions of relatively high sequence similarity, this

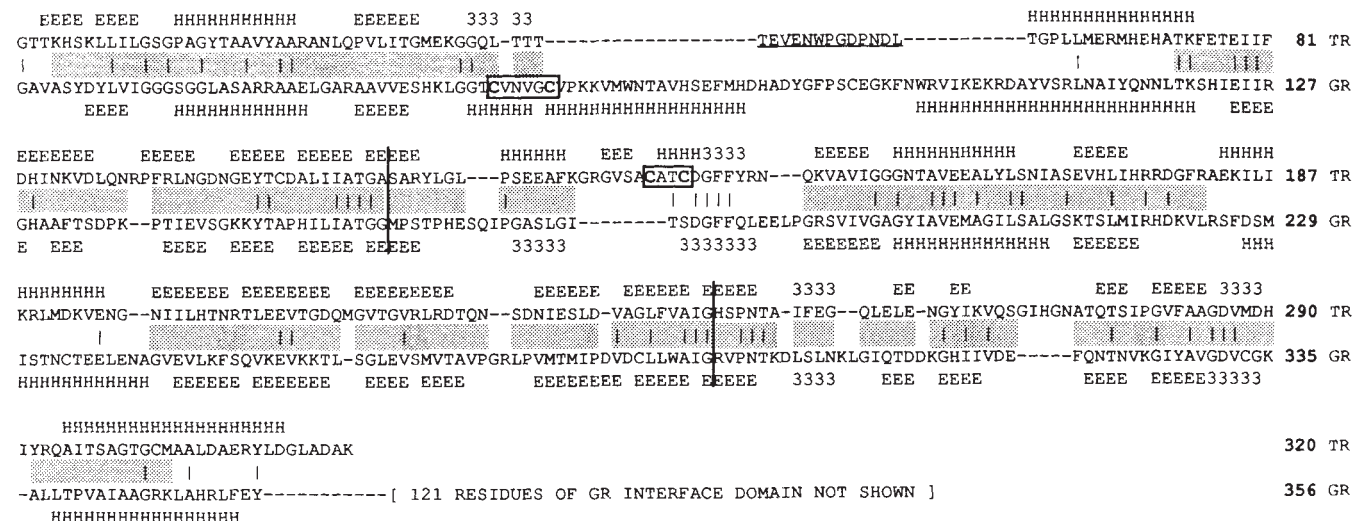


FIG. 3 Sequence alignment of *E. coli* TR⁶ and human erythrocyte GR¹ based on secondary and tertiary structure (single-letter amino-acid notation). A less extensive but similar alignment based on sequence alone has been reported⁷. Boundaries between the FAD, central and NADPH domains are shown by vertical lines; the N- and C-terminal segments constitute the FAD and central domains, with the NADPH domain in the middle. Vertical bars indicate sequence identity, and the active-site cysteines are boxed. The overall sequence identity is 21%. Secondary structure code: E, β sheet; H, α helix; 3, 3_{10} helix; the assignment for TR was done on the basis of hydrogen-bonding patterns²⁴, with somewhat relaxed criteria as described for GR¹. Regions of tertiary structural similarity are shaded. The FAD/central and the NADPH domains are treated separately for the evaluation of this similarity, as their relative orientations differ (Fig. 2). A least-squares procedure was used to superimpose C $_{\alpha}$ atoms, initially by aligning the secondary structure elements. Insertions and deletions in the sequences

were then determined by increasing the list of residue pairs used in the least-squares superpositioning and removing those that appeared to be structurally inequivalent. Although subjective, the results are unambiguous in most cases. Residues 47–59 of TR are underlined to indicate that they do not align with the residues below. Only four aligned segments are not considered to be structurally equivalent. All are helices that are present in both structures but play unique roles in TR and have changed their orientations substantially: residues 60–72, 183–195 and 305–320 (TR) are part of the dimer interface, and residues 135–14 include the redox-active cysteines. 137 residues in the FAD/central domains and 96 in the NADPH domain are considered structurally equivalent. This corresponds to 72% and 74% of the total in each part of the TR sequence, respectively, and the root-mean-square deviation from equivalent residues in GR is 1.8 Å and 2.0 Å, respectively.

suggests that both proteins have evolved from a common ancestor. The FAD and NADPH domains of TR and GR both contain the $\beta\alpha\beta$ nucleotide-binding motif⁹ and an additional anti-parallel β sheet, and may have arisen by gene duplication¹⁰. Previous comparison of GR and an FAD-containing monooxygenase, *p*-hydroxybenzoate hydroxylase, revealed structural similarity in their FAD domains and some common elements in their central domains¹¹. Elements of divergent as well as convergent evolution were indicated, but distinguishing between the two mechanisms was difficult because of the absence of sequence identity, the lack of structural similarity in the NADPH

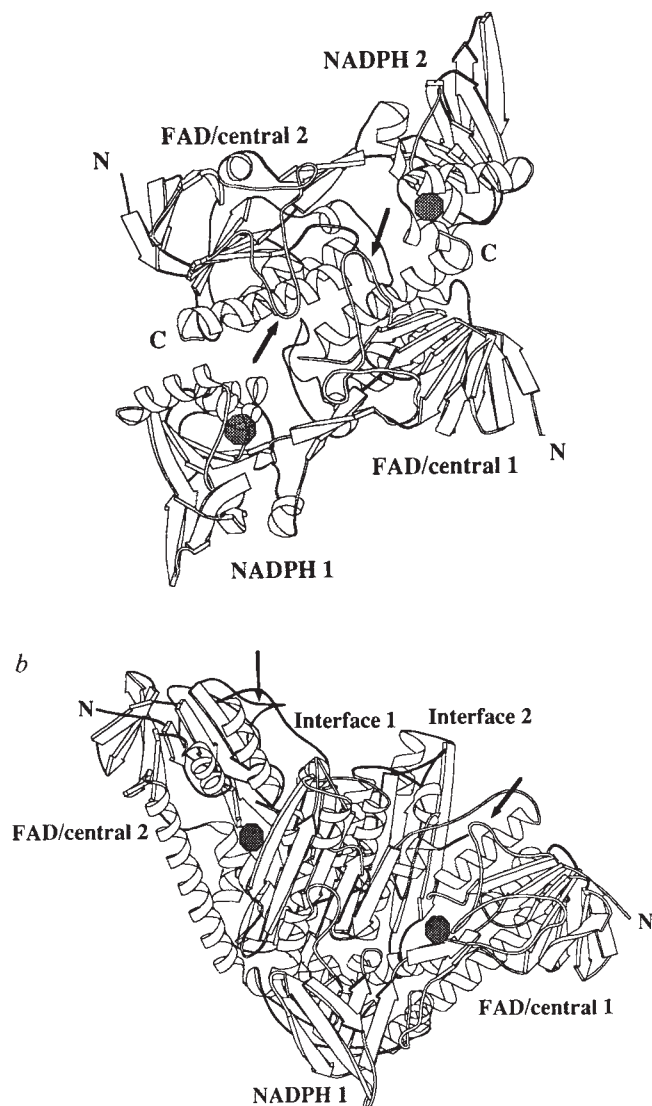


FIG. 4 Schematic diagrams of the dimeric structures of *a*, TR and *b*, GR (ref. 1). Helices are shown by ribbons, strands of β sheet by arrows²⁵. Domains in different monomers within the dimers are labelled 1 and 2 and the FAD domains of monomer 1 are in the same orientation. One of the NADPH domains of GR is hidden. The approximate locations of the redox-active cysteines are shown by shaded circles. The interface domains, shown here for GR, have no counterpart in TR. The arrows indicate two symmetry related loops that in TR interact across the dimer interface and are widely separated in GR. The change in relative orientation of these loops indicates the large change in relative orientation of the monomers. The total surface area buried on dimer formation in TR is 2,950 Å², of which 39%, 44% and 17% are contributed by the FAD, central and NADPH domains, respectively. This is in contrast to GR, where the buried surface area is much larger (3,590 Å²), and where the interface domain provides the largest buried surface area (57%), along with a smaller contribution¹ from the FAD domain (37%). The accessible surface areas were calculated using a probe of radius 1.6 Å (ref. 26).

domains and fundamental differences in the reactions catalysed¹¹. By contrast, TR and GR catalyse very similar reactions that differ only in substrate, and share a sequence similarity showing relationships in all three domains of TR⁶.

In contrast to the similarity in tertiary structure, the dimeric forms of the enzyme are completely different (Fig. 4). Dimer formation in GR involves an interlocking of the interface domains, with each one making contacts with the interface and FAD domains of the other molecule and forming the deep crevices that are the binding sites for glutathione¹. In TR, there are two shallow depressions at the dimer interface, on the same face of the enzyme. These are probably the thioredoxin binding sites, as the side chains of both reactive cysteines are exposed at each depression consistent with mechanistic studies¹². Dimer formation in TR requires no additional structural elements other than those that are also present in GR (Fig. 4). The conformation of the FAD group and its position in the FAD/central domains is conserved between TR and GR (Fig. 2). Relative to the flavin ring system, the other components of the catalytic machinery are considerably rearranged: the redox-active disulphides are on opposite sides of the flavin ring system in the two enzymes, and the NADPH domains are rotated with respect to one another. Note that the redox active cysteines in TR stack against the flavin (Fig. 1) in a location that is coincident with that of the nicotinamide ring of NADPH in GR¹³, indicating that the mode of pyridine nucleotide binding in TR is very different. In GR the redox-active cysteines and the catalytic base (His 467) are contributed by different monomers¹, whereas in TR the probable active site base (His 245)¹⁴ is in the same molecule as the cysteines.

The process of convergent evolution that has led to the alternative dimeric structures and active sites has occurred with the preservation of the structural framework of the FAD/central and NADPH domains, confirming the intrinsically modular construction of these enzymes. Another example of the independent acquisition of similar function by proteins that have diverged from a common ancestor is the dimeric clam haemoglobin, which has subunits that resemble those of tetrameric mammalian haemoglobin¹⁵. The dimer interface is completely different, however, and provides for an allosteric mechanism which is unrelated to that found in mammalian haemoglobins¹⁵. □

Received 21 March; accepted 10 May 1991.

- Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **195**, 701–729 (1987).
- Williams, C. H. Jr *The Enzymes* 3rd edn **13**, 89–173 (1976).
- Schirmer, R. H. & Schulz, G. E. in *Pyridine Nucleotide Coenzymes Part B (Coenzymes and Cofactors)* Vol. 2 (eds Dolphin, D., Poulson, R. & Avramovic, O. 333–379 (Wiley New York, 1987).
- Thieme, R., Pai, E. F., Schirmer, R. H. & Schulz, G. E. *J. molec. Biol.* **152**, 763–782 (1981).
- Russel, M. & Model, P. *J. biol. Chem.* **263**, 9015–9019 (1988).
- Williams, C. H. Jr., Prongay, A. J., Lennon, B. W. & Kuriyan, J. in *Flavins and Flavoproteins* (eds Curti, B., Zannetti, G. & Ronchi, S.) (de Gruyter, Berlin, in the press).
- Holmgren, A. *J. biol. Chem.* **264**, 13963–13966 (1989).
- Chothia, C. & Lesk, A. M. *EMBO J.* **5**, 823–826 (1986).
- Rossmann, M. G., A. Liljas, C. I. Brändén & Benasak, L. J. in *The Enzymes* (ed. Boyer, P. D.) 61–102 (Academic, New York, 1975).
- Schulz, G. E. *J. molec. Biol.* **145**, 335–347 (1980).
- Wierenga, R. K., Drenth, J. & Schulz, G. E. *J. molec. Biol.* **167**, 725–739 (1983).
- O'Donnell, M. E. & Williams, C. H. Jr *Biochemistry* **24**, 7617–7621 (1985).
- Karplus, P. A. & Schulz, G. E. *J. molec. Biol.* **210**, 163–180 (1989).
- O'Donnell, M. E. & Williams, C. H. Jr *J. biol. Chem.* **258**, 13795–13805 (1983).
- Roy, W. E. J., Hendrickson, W. A. & Chiancone, E. *Science* **249**, 518–521 (1990).
- Prongay, A. J., Engelke, D. R. & Williams, C. H. Jr *J. biol. Chem.* **264**, 2656–2664 (1989).
- Kuriyan, J., Wong, L., Russel, M. & Model, P. *J. biol. Chem.* **264**, 12752–12753 (1989).
- Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39**, 813–817 (1983).
- Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Brünger, A. T. *X-PLOR (Version 1.5) Manual* (The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, Connecticut, 1988).
- Jones, T. A. & Thirup, S. *EMBO J.* **5**, 819–822 (1986).
- Ramachandran, G. N. & Sasisekharan, V. *Adv. Protein Chem.* **23**, 283–437 (1968).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
- Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
- Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).

ACKNOWLEDGEMENTS. We thank W. A. Hendrickson, J. W. Pflugrath, R. M. Sweet and G. A. Petsko for encouragement, advice and use of their X-ray equipment. Beamline X12-C at the National Synchrotron Light Source at Brookhaven National Laboratory is funded by the Office of Health and Environmental Research of the US Department of Energy. This work was supported by the NIH and the NSF. J.K. is a Pew Scholar in the Biomedical Sciences.